

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 18, 2001, 03:18:53 ; Search time 3860.4 Seconds
(without alignments)
8809.648 Million cell updates/sec

Title: US-09-434-382-1
Perfect score: 2481
Sequence: 1 atgtgggctttgtctgct.....agaagtcagagccagtgta 2481

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1118133 seqs, 6853842396 residues 2236266
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

1: gb_bal.*
2: gb_bal2.*
3: gb_om.*
4: gb_ov.*
5: gb_ph.*
6: gb_pl1.*
7: gb_pl2.*
8: gb_pr1.*
9: gb_pr2.*
10: gb_pr3.*
11: gb_ro.*
12: gb_sy.*
13: gb_un.*
14: em_fun.*
15: em_hum1.*
16: em_hum2.*
17: em_in.*
18: em_om.*
19: em_or.*
20: em_ov.*
21: em_pat.*
22: em_ph.*
23: em_pl.*
24: em_ro.*
25: em_sts.*
26: em_sy.*
27: em_un.*
28: em_vi.*
29: gb_htg1.*
30: gb_htg2.*
31: gb_in1.*
32: gb_in2.*
33: gb_bal1.*
34: gb_bal2.*
35: em_hum3.*
36: em_hum4.*
37: gb_pr4.*
38: gb_htg3.*
39: gb_htg4.*
40: gb_htg5.*
41: gb_htg6.*
42: gb_htg7.*
43: em_htg1.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	ID	Description
1	2448.2	98.7	2976	37 AK001392	AK001392 Homo sapi
2	247.4	10.0	118788	8 AC005277	AC005277 Homo sapi
3	133.4	5.4	2690	32 AF215894	AF215894 Drosophil
4	133.4	5.4	115873	41 AC017383	AC017383 Drosophil
5	133.4	5.4	125448	29 AC007417	AC007417 Drosophil
6	133.4	5.4	190574	29 AC007352	AC007352 Drosophil
7	133.4	5.4	261846	31 AE003830	AE003830 Drosophil
8	84.2	3.4	31201	47 SPAC1D4	269239 S.pombe chr
9	71.8	2.9	46335	6 AC022354	AC022354 Arabidops
10	70	2.8	19158	6 AF188714	AF188714 Emericell
11	67	2.7	32329	47 SPBC3D6	295620 S.pombe chr
12	65	2.6	143209	7 AP001550	AP001550 Oryza sat
13	58.8	2.4	75289	6 AB023046	AB023046 Arabidops
14	58.4	2.4	3685	47 SCYKR079C	228304 S.cerevisia
15	56.8	2.3	23533	1 AE001039	AE001039 Archaeogl
16	53.8	2.2	100005	65 AC079030	AC079030 Homo sapi
17	53.8	2.2	145440	52 AC026119	AC026119 Homo sapi
18	53.8	2.2	180260	57 AC068797	AC068797 Homo sapi
19	52.4	2.1	12026	1 AE005109	AE005109 Halobacte
20	51.8	2.1	11836	1 AE000936	AE000936 Methanoba
21	51.4	2.1	729	78 HSY18896	Y18896 Homo sapien

Db 675 GAAAGGCTCTCAGAGGCTCAGTCCAGAGCGATCTTCAGACTCGAGTGAATCAAAAT 734
Qy 661 gagccacacettccacatgggttagccagagaagagggtcagggaactcttccctggtc 720
Db 735 GAGCACACCTTCCACATGTTGTTAGCCAGAGAGAGGGGTTCAGGGACTCTTCCTGGTC 794
Qy 721 gtacgtttcatctgaagcttcacattaaagagagaaactcttgggtcctcaagcaaaag 780
Db 795 GTAGCTTCACTGTAGCTTCACTTAAAGAGAGAAATCTTGGTGGTCTCAAGCAAAAG 854
Qy 781 gagatgggctccagttgggacagctgcccacgtctcccatcattgctgctgctcaagac 840
Db 855 GAGATGGGCTCCACAGTTGGGACAGCTGCCATCGCTCCCATCATTTGCTGCTCAAGGAC 914
Qy 841 gggaaaagcatcactcatgaaggagagagattttggctgaagagctggtactcctcca 900
Db 915 GGGAAAAGCATCACTATGAAGAGAGAGAGATTTTGGTGAAGAGCTGTGTACTCTCTCCA 974
Qy 901 gatcctgggtgctgtttgttggtgagaatgtccagatgaagagcttcaatcaaccatc 960
Db 975 GATCCTGGTGTGCTTTTGTGGTGGTGAATGTCCAGATGAAGCTTTCATTCAACCCCATC 1034
Qy 961 tgtgagatgccactttcagaggtaccaggaagagcagatgcccccggtggccttgggtg 1020
Db 1035 TGTGAGATGCCACTTTCAGAGTACCAGGAAGGCGAGATGCCCCCTGGCTTTGGTG 1094
Qy 1021 gttcacatggccccagcatctgtgttgacagcaggtaccagcagtgatgagagagg 1080
Db 1095 GTTCACATGGCCCCAGCATCTGTCTTGTGGACAGCAGGTACCAGCAGTGGAGAGG 1154
Qy 1081 tttgggctgacacccagcaacttgctcctgaatgagaaactgtgctcagtttcaaacctt 1140
Db 1155 TTTGGGCTGTACACCCAGCACTTGTGCTGAATGAGAACTGTGCTCAGTTTCACAACTT 1214
Qy 1141 cgcagcacaaagttaaaacccagctcaacctcaaccgcgacatcttccccctgctc 1200
Db 1215 CCGAGCCACAGATTCAAACCCAGCTCAACCTCATCCACCCGGACATCTTCCCTCTGCTC 1274
Qy 1201 accagtttccgctgaagaagaggggccccaccctcagtgcccaatggttcagggtgaa 1260
Db 1275 ACCAGTTTCCGCTGT---AAGGAGGGCCCCACCCTCAGTGTGCCATGTTTCAGGGTGA 1331
Qy 1261 tgcctcctaagtaaccagctccgtcccaaggaggagtgccagagggatgcattattact 1320
Db 1332 TGCTCTCTCAAGTACCAGTCCGCTCCAGAGGAGTGGCAGAGGGATGCCATTATTACT 1391
Qy 1321 tgaatcctgaggaatcattagttgagcgtcagcttcccaacttccacaaagcgtg 1380
Db 1392 TGCATCTCTGAGGAATTCATAGTTGAGCGCTGCAGCTTCCCACTTCCAGCAGAGCGTG 1451
Qy 1381 caggagtacaggaggagtgccagagagggccccagccccagcagagaaagagtcagtac 1440
Db 1452 CAGGAGTACAGGAGAGTGGCAGAGCGGCCGCCGCCAGCAGAGAAAGTCACTAC 1511
Qy 1441 ccagaaatcattctccttggaacgggtctgtccatcccgatgaagattcgaatgtcagt 1500
Db 1512 CCAGAAATCATCTCTCTTGGAAACAGGCTGTGCCATCCGATCCGATGAAGATTGCAATGTCAGT 1571
Qy 1501 gccacactgtcaacataagcccgacacgtctctgtactgagactggtgagggcaca 1560
Db 1572 CCCACACTTGTCAACATTAAGCCCCGACACGCTCTGTCTACTTGGACTGTGGTGGGACACA 1631
Qy 1561 ttgggcagctgtgcgtcattacggagacaggtggagaggttccctgggcacccctggct 1620
Db 1632 TTTGGGACAGCTGTGCCGTCATTACGGAGACAGGTGGACAGGGTCTTGGGACCCCTGGCT 1691
Qy 1621 gctgtgttgggttcccaactgacgcagatcacacacgggcttgccaaagtatttggctg 1680
Db 1692 GCTGTGTGTGTGCTCCACCTGCAGCGCATCACACACGGGCTTGCCAAATATCTTGCTG 1751
Qy 1681 cagagagacgcgcttggatctttgggaagcggcttcaacctttgtcgtggtggtgcc 1740
Db 1752 CAGAGAGAACCGGCTTGGCAATCTTTGGGAAGCGGCTTCAACCCCTTGTGCTGGTGGTCC 1811

Qy 1741 cccaaccagctcaaaagcctggctccagcaggtaccacaaccaggtccagagaggtcctgcac 1800
Db 1812 CCCAACCAAGCTCAAAAGCCTGGCTCCAGCAGTACTAAACCAAGTGCAGAGGTCCTGCAC 1871
Qy 1801 cacatcagtatgattcctgcctcaaatgcttcaggaagggctgagatctccagctcgtca 1860
Db 1872 CACATCAGTATGATTCCTGCCAAATGCCCTTCAGGAAGGGCTGAGATCTCCAGTCTCGCA 1931
Qy 1861 gtggaagatgatcagttcgtgttgcaacatgatttggaagagtttcagaccctgt 1920
Db 1932 GTGGAAAGATTGATCAGTTCGCTGTTCGGAACATGTGATTTGGAAGAGTTTCAGACCTGT 1991
Qy 1921 ctggtgcggcactgcaagcatgctgttgctgtgctgctggtgcacacctctcgtgctggaaa 1980
Db 1992 CTGTTGCGGCACCTGCAAGCATGCGTTTGGCTGTGCGCTGGTGCACACCTCTGGCTGGAAA 2051
Qy 1981 gtggtctattccgggggacacacatgcccctgcagaggtctggtccggatggggaagatgcc 2040
Db 2052 GTGGTCTATTCCGGGGACACCATGCCCTCCGAGGCTGTGGTCCGATGGGGAAGATGCC 2111
Qy 2041 accctcctgatcatgaaagccacctggaagatggtttggaagagaaagcagctggaag 2100
Db 2112 ACCCTCTGTATACATCAAGCCACCTCGAAGATGTTTGGAAAGAGACAGTGGAAAAAG 2171
Qy 2101 acacacagcaaacactcccaagccatcagctggtgggatgctggaatgaacgcggagtctatt 2160
Db 2172 ACACACAGCAACAACGTCCTCAAGCCATCAGCTGGGGATGCGGATGAACGCGGAGTTCATT 2231
Qy 2161 atgctgaaccacttcagcagcagctatgccaaggtccccctcttcagcccccaacttcagc 2220
Db 2232 ATGCTGAACCACTTTCAGCAGCGCTATGCCAAGGTCCCCCTCTTCAGCCCCAACCTCAGC 2291
Qy 2221 gagaaagtggaggttgctcttgacacacatgaaggtctgctttggagactttccacaatg 2280
Db 2292 GAGAAAGTGGAGTGGCTTTGACCACATGAAGCTCTGCTTGGAGACTTTCACACATG 2351
Qy 2281 cccaagctgattccccccactgaaagccctgttctgctgacatcagagagatgagaggag 2340
Db 2352 CCCAAGCTGATTCCCCCACTGAAGCCCTGTTTGTGGCGACATCGAGGAGATGGAGGAG 2411
Qy 2341 cgcagggagagcggggagctgcggcaggtgcggcgcctcctctccagagagctggca 2400
Db 2412 CGCAGGGAGAGCGGGAGCTGCGGAGGTGCGGGCGGCCCTCTCTTCCAGGGAGTGGCA 2471
Qy 2401 ggcgccttgaggtgagagcctcagcagaaagcggccccacacagagagccacagcgc 2460
Db 2472 GCGGCTTGGAGGAT-GGGAGCTCAGCAGAAAGCGGGCCCCACAGAGGACACAGGCC 2530
Qy 2461 aagaaggtcagagcccgagtga 2481
Db 2531 AAGAAGGTCAAGCCCGAGTGA 2551

RESULT 2

AC005277
LOCUS AC005277 118788 bp DNA PRI 23-JUL-1998
DEFINITION Homo sapiens chromosome 17, clone hRPK.597_M.12, complete sequence.
ACCESSION AC005277
VERSION AC005277.1 GI:3337311
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 118788)
AUTHORS Birren,B., Fasman,K., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 17, clone hRPK.597_M.12
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 118788)
AUTHORS Birren,B., Fasman,K., Linton,L., Nusbaum,C., Lander,E., Allen,N.,
Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Boatn,C.,
Boutwell,C., Brown,A., Castle,A., Cerny,J., Cooke,P., Depayre,E.,

Devon, K., Dewar, K., Donelan, L., Etemadi, S., Ferreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D., Gardyna, S., Gensheimer, S., Geraghty, K., Gilmartin, T., Grant, G., Hagos, B., Harris, K., Horton, L., Howland, J.C., Hui, L., Jacotot, L., Kann, L., Macdonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K., Meldrum, J., Molla, M., Morris, W., Morrow, J., Mychaleckyj, J., Nachman, A., Nahf, R., Naylor, J., Niloff, M., O'Connor, T., Pavlin, B., Peterson, K., Riley, R., Roberts, D., Rossello, R., Roy, A., Shyam, R., Stange-Thomann, N., Stilwell, J., Stojanovic, N., Stone, C., Strickland, C., Subramanian, A., Torruella-Miller, I., Vassiliev, H., Vo, A., Wagner, A., Wang, B., Wheeler, J., Wu, Y., Ye, W.J., Zhao, J. and Zody, M.

TITLE

JOURNAL Submitted (10-JUL-1998) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE 3 (bases 1 to 118788)

AUTHORS

Birren, B., Fasman, K., Linton, L., Nusbaum, C., Lander, E., Allen, N., Baker, J., Baldwin, J., Barna, N., Beckerly, R., Benn, J., Boatin, C., Boutwell, C., Brown, A., Castle, A., Cerny, J., Cooke, P., Depayre, E., Devon, K., Dewar, K., Donelan, L., Etemadi, S., Ferreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D., Gardyna, S., Gensheimer, S., Geraghty, K., Gilmartin, T., Grant, G., Hagos, B., Harris, K., Horton, L., Howland, J.C., Hui, L., Jacotot, L., Kann, L., Macdonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K., Meldrum, J., Molla, M., Morris, W., Morrow, J., Mychaleckyj, J., Nachman, A., Nahf, R., Naylor, J., Niloff, M., O'Connor, T., Pavlin, B., Peterson, K., Riley, R., Roberts, D., Rossello, R., Roy, A., Shyam, R., Stange-Thomann, N., Stilwell, J., Stojanovic, N., Stone, C., Strickland, C., Subramanian, A., Torruella-Miller, I., Vassiliev, H., Vo, A., Wagner, A., Wang, B., Wheeler, J., Wu, Y., Ye, W.J., Zhao, J. and Zody, M.

TITLE

JOURNAL Submitted (23-JUL-1998) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT On Jul 23, 1998 this sequence version replaced gi:3335015.

All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Only the first 118.8 kilobases of this clone are being submitted. The remainder overlaps accession number AC005274 (WICGR project L350).

FEATURES

source

Location/Qualifiers

1. .118788
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="hRPK 597_M.12"
/clone_lib="RPCI-11 human BAC library"
/map="17"
/chromosome="17"

repeat_region 795..884
repeat_region /rpt_family="MIR"
repeat_region complement(932..1000)
repeat_region /rpt_family="MIR"
repeat_region 2044..2076
repeat_region /rpt_family="AT_rich"
repeat_region 2237..2713
repeat_region /rpt_family="MER34"
repeat_region 2718..3477
repeat_region /rpt_family="L2"
repeat_region complement(3478..3821)
repeat_region /rpt_family="MLT1A1"
repeat_region 3822..4007
repeat_region /rpt_family="L2"
repeat_region complement(4015..4281)
repeat_region /rpt_family="MER89"
repeat_region complement(4282..4321)
repeat_region /rpt_family="(TAAA)n"
repeat_region complement(4328..4608)
repeat_region /rpt_family="AluSx"
repeat_region complement(4609..4895)
repeat_region /rpt_family="MER89"
repeat_region complement(4896..5186)

repeat_region /rpt_family="AluSp"
5219..5479
repeat_region /rpt_family="L2"
complement(5481..5740)
repeat_region /rpt_family="MER34"
complement(5963..6275)
repeat_region /rpt_family="AluSp"
complement(6411..6494)
repeat_region /rpt_family="MER5A"
6830..7013
repeat_region /rpt_family="MLT1C"
7014..7318
repeat_region /rpt_family="AluYb8"
7319..7581
repeat_region /rpt_family="MLT1C"
7721..8032
repeat_region /rpt_family="AluJb"
8039..8333
repeat_region /rpt_family="AluSq"
9431..9460
repeat_region /rpt_family="AT_rich"
9662..9829
repeat_region /rpt_family="MIR"
complement(11302..11728)
repeat_region /rpt_family="MER39"
complement(11824..11884)
repeat_region /rpt_family="LTR16C"
complement(11928..12108)
repeat_region /rpt_family="LTR16C"
12453..12725
repeat_region /rpt_family="AluJb"
13377..13785
repeat_region /rpt_family="MSTB"
14118..14153
repeat_region /rpt_family="(CA)n"
complement(15320..15620)
repeat_region /rpt_family="AluJo"
15683..15759
repeat_region /rpt_family="(CA)n"
complement(15838..15893)
repeat_region /rpt_family="MER91A"
complement(15915..16080)
repeat_region /rpt_family="MER91A"
complement(16474..16544)
repeat_region /rpt_family="L2"
17907..17920
repeat_region /rpt_family="LIME"
18271..18311
repeat_region /rpt_family="L2"
complement(18368..18529)
repeat_region /rpt_family="AluJb"
complement(18530..18560)
repeat_region /rpt_family="(CA)n"
complement(18561..18681)
repeat_region /rpt_family="AluJb"
18760..19065
repeat_region /rpt_family="AluSg"
complement(19638..19896)
repeat_region /rpt_family="MIR"
complement(19971..20260)
repeat_region /rpt_family="LIME3A"
20263..20552
repeat_region /rpt_family="MER58B"
complement(20553..20816)
repeat_region /rpt_family="LIME3A"
complement(20821..21069)
repeat_region /rpt_family="LIM4"
complement(21071..21100)
repeat_region /rpt_family="(TAAA)n"
complement(21101..21338)
repeat_region /rpt_family="AluSx"
complement(21348..21598)
repeat_region /rpt_family="LIM4"

Db 1744 CTGGCTCTCGCCAAATTAACCTTGGTT---GGAGTTCTACAATCGACAATAGAAACC 1800
 Qy 1795 ctgcacacatcagtagtattcctgcgaatgccttcaggagggcgtgagatctccagt 1854
 Db 1801 GTAGAAGATCGGTATACCTAGTGGCAACGGTGAACCTCTTGTCTAGTCCCTGAGCGGT 1860
 Qy 1855 cctcagtggaagattgatcagtcgtctgttggaacatgtatttggaagagttcag 1914
 Db 1861 GAACAAGTGTGAACGTAGGAATACGTCATATCCACCTGCTAGTTAGGCACTGCCCC 1920
 Qy 1915 acctgtctgtggtgcactgaacatgcgttttggtctgtgctgtggtgcacacctctggc 1974
 Db 1921 AACCTCTTTCGGAATAAGCCTTAACCTTGCGC-----GCAAAAGCAACAATAGCGAAACC 1971
 Qy 1975 tggaaagtggtctattccggtggaacaccatgccttcgagagctctggtccggtgagggaaa 2034
 Db 1972 GTCAAGATCACTTACAGCGGGGACACTATGCCATGTCAAGGATCTGATCGATCTGGCGCCG 2031
 Qy 2035 gatccaccctctctgatacatgaagccaccctggaagatggtttggaagaggaagcagt 2094
 Db 2032 GACTCCACCGTCTTATTCACGAGCGCAATGGAGGACGACTTGGAGGAGGAGGCGCGG 2091
 Qy 2095 gaaagacacacagcaaacgtcccaagccaatcagcgtggggatgctggaacagcggag 2154
 Db 2092 TTAAGACGCAACAGCGCTGTCAGAGCTATTTCAGCAGGGGGCGGAATATGAACGCCCGC 2151
 Qy 2155 ttcattatgctgaaccacttcagcagcgtatgcgaaggtcccccctt---cagcccc 2211
 Db 2152 CATAGATTCTACCCACTTTTCGAGGGTTACGCTAAGTCCCGCGGATTACCGAGTGAT 2211
 Qy 2212 aactcagcggagaagtggtggttccttgaccacatgaaggtctgctttggagacttt 2271
 Db 2212 GAGGATATGCAACGAGTGTCAATAGCCTTTTGATATATGAAGTGACCATTTGAGGATCTG 2271
 Qy 2272 ccaacaatgcccagctgattccccactgaagccctgtttgtgctggtggaacatcagagag 2331
 Db 2272 CAGCACTACCAACAGCTCTACCCCGCCCTTTTCGCGATGTACGGCGAGTACACGGGAAGA 2331
 Qy 2332 atggagagcagcaggaggaagcggagctgcgaggtgctg 2372
 Db 2332 CTGGAGCAGCGGCTGTCAACGGGAGCTGAACAGGAGCG 2372

RESULT 4
 AC017383 115873 bp DNA HTG 09-DEC-1999
 LOCUS Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
 DEFINITION pieces.
 ACCESSION AC017383
 VERSION AC017383.1 GI:6553603
 KEYWORDS HTG; HTGS_PHASE2.
 SOURCE fruit fly.
 ORGANISM Drosophila melanogaster
 Eukaryote; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 115873)
 AUTHORS Adams, M. and Venter, J.C.
 TITLE Direct Submission
 JOURNAL Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
 Rockville, MD, USA
 COMMENT This sequence was identified as CDM:10210513 by the submitter.
 For more information on this record e-mail to fly@celera.com.
 * NOTE: This is a 'working draft' sequence.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.

FEATURES
 source
 Location/Qualifiers
 1..115873
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 BASE COUNT 31582 a 25315 c 25855 g 33121 t

ORIGIN

Query Match 5.4%; Score 133.4; DB 41; Length 115873;
 Best Local Similarity 50.5%; Pred. No. 6.5e-20;
 Matches 475; Conservative 0; Mismatches 436; Indels 30; Gaps 5;

Qy 1438 taccacgaatacatctctcttggaacagggcttcgcatcccgatgaagattcgaatgctc 1497
 Db 63659 TACCCGAAGATAATTTCTTGGGCACAGCTCTCTCATTTCCCAATAAGACGGAACGTG 63718
 Qy 1498 agtgcacacattgtcaacataagccccgacacgtctctgtactactgaactgtggtgagggc 1557
 Db 63719 AGCTCCATTCTGATCAGGACTGCAATTCATGCTACGTCTGTTGTTGGATTGTGGAAGGA 63778
 Qy 1558 acatttggcagcgtgcgctcattacggagaccag---gtggacagggctcctgggcacc 1614
 Db 63779 ACTTATGTCATAATTTGTACGACTATATGCACACGAAAGGCGCATTTGATTTCTGCCAG 63838
 Qy 1615 ctggctgctgtgttctccaccctgcacgagatccaccacacggcgttgcgaagtatc 1674
 Db 63839 CTACAGGCGATTTATGATATCCATTTGATGCCGACCAACATTTGGATTGGATTGCTGTG 63898
 Qy 1675 ttgctgcagagagaacgcgcttggcatcttcttggaagacgcttccaccttctgctggtg 1734
 Db 63899 CTCCGGGCAAG-----GAGGCAATTAACACCTAGGGCAGACCCACTTATTCTG 63946
 Qy 1735 gttcccccacacagctcaaaagcctgcctcagcagtaaccacacacagtgccagaggtc 1794
 Db 63947 CTGGCTCTCGCCAAATTTGAACCTTTGGTT---GGAGTTCTACAATCGACAATAAGAACCC 64003
 Qy 1795 ctgcacacatcagtagtattcctgcgaatgccttcaggagggcgtgagatctccagt 1854
 Db 64004 GTAGAAGATCGTATACCTAGTGGCAACGTTGAACCTCTAGCTAGTCTCCCTGAGCGGT 64063
 Qy 1855 cctcagtggaagattgatcagttcgtgttggaacatgtgatttggagaggttccag 1914
 Db 64064 GAACAAGTGTGAACGTTCTTATTCACGAGCGCAATGGAGGATGACTTTGGAGGAGGCGCGG 64123
 Qy 1915 acctgtctgtgctgcgacatgcgaacatgccttggctgctgctgctgcacacctctggc 1974
 Db 64124 AACTCTTTCCGAATAAAGCCTTAACCTTGCGC-----GCAAAAGCAACAATAGCGAACCC 64174
 Qy 1975 tggaaagtggtctattccggtggaacacatgccttcgagaggtctctgctcggatgggaaa 2034
 Db 64175 GTCAAGATCACTTACAGCGGCGACACTATGCCATGTCAAGGATCTGATCGATCTGGCGCGC 64234
 Qy 2035 gatccaccctctctgatacatgaagccaccctggaagatggtttggaagaggaagcagt 2094
 Db 64235 GACTCCACCGTCTTATTCACGAGCGCAATGGAGGATGACTTTGGAGGAGGAGGCGCGG 64294
 Qy 2095 gaaagacacacagcaaacgtcccaagccaatcagcgtggggatgctggaacagcggag 2154
 Db 64295 TTAAGACGCAACAGCGCTGTCAGAGCTATTTCAGCAGGGGGCGGAATATGAACGCCCGC 64354
 Qy 2155 ttcattatgctgaaccacttcagcagcgtatgcgaaggtcccccctt---cagcccc 2211
 Db 64355 CATAGATTCTCTCCCACTTTTCGAGCGGTTACGCTAAGTCCCGCGGATTACCGAGTGAT 64414
 Qy 2212 aactcagcggagaagtggtggttccttgaccacatgaaggtctgctttggagacttt 2271
 Db 64415 GAGGATATGCAACGAGTGTCAATAGCCTTTTGATATATGAAGTGACCGTTGAGGATCTG 64474
 Qy 2272 ccaacaatgcccagctgattccccactgaagccctgtttgtcgtggtggaacatcagagag 2331
 Db 64475 CAGCACTACCAACAGCTCTACCCCGCCCTTTTCGCGATGTACGGCGAGTACACGGGAAGA 64534
 Qy 2332 atggagagcagcaggaggaagcggagctgcgaggtgctg 2372
 Db 64535 CTGGAGCAGCGGCTGTCAAGCGGAGCTGAACAGGAGCG 64575

RESULT 5

AC007417/c

LOCUS

DEFINITION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

AC007417 125448 bp DNA HTG 02-AUG-1999
 Drosophila melanogaster chromosome 2 clone BACR48F07 (D625) RPCI-98
 48.F.7 map 47A-47B strain Y; cn bw sp, *** SEQUENCING IN PROGRESS
 ***, 94 unordered pieces.

AC007417
 AC007417.3 GI:5670592
 HTG; HTGS_PHASE1.
 fruit fly.

Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 125448)
 Celisner, S.E., Agbayan, A., Arcaina, T.T., Baxter, E., Blazej, R.G.,
 Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,
 Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,
 Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L.,
 Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P.,
 Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S.,
 Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Snir, E.,
 Svirskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and
 Rubin, G.M.

Sequencing of Drosophila melanogaster
 Unpublished
 2 (bases 1 to 125448)
 Celisner, S.E., Agbayan, A., Arcaina, T.T., Baxter, E., Blazej, R.G.,
 Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,
 Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,
 Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L.,
 Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P.,
 Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S.,
 Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Snir, E.,
 Svirskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and
 Rubin, G.M.

Direct Submission
 Submitted (28-Apr-1999) Drosophila Genome Center, Lawrence Berkeley
 Laboratory, MS 64-121, Berkeley, CA 94720, USA
 On Aug 2, 1999 this sequence version replaced gi:5629944.
 For further information about this sequence, including its location
 and relationship to other sequences, please visit our sequence
 archive Web site (<http://www.fruitfly.org/sequence/>) or send email
 to bdg@fruitfly.berkeley.edu. All contigs in this submission meet
 the following cutoffs: length >= 200 bases.
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 94 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 969: contig of 969 bp in length
 970 1049: gap of unknown length
 1050 2078: contig of 1029 bp in length
 2079 2158: gap of unknown length
 2159 3249: contig of 1091 bp in length
 3250 3329: gap of unknown length
 3330 4547: contig of 1218 bp in length
 4548 4627: gap of unknown length
 4628 5262: contig of 635 bp in length
 5263 5342: gap of unknown length
 5343 6096: contig of 754 bp in length
 6097 6176: gap of unknown length
 6177 7486: contig of 1310 bp in length
 7487 7566: gap of unknown length
 7567 8160: contig of 594 bp in length
 8161 8240: gap of unknown length
 8241 9090: contig of 850 bp in length
 9091 9170: gap of unknown length
 9171 9503: contig of 733 bp in length
 9504 9984: gap of unknown length
 10564: contig of 581 bp in length

10644: gap of unknown length
 11655: contig of 1011 bp in length
 11735: gap of unknown length
 12539: contig of 724 bp in length
 12539: gap of unknown length
 13197: contig of 658 bp in length
 13277: gap of unknown length
 14315: gap of unknown length
 14394: gap of unknown length
 15048: contig of 654 bp in length
 15128: gap of unknown length
 16091: contig of 963 bp in length
 16171: gap of unknown length
 16817: contig of 646 bp in length
 17598: gap of unknown length
 17598: contig of 701 bp in length
 17598: gap of unknown length
 18750: contig of 1072 bp in length
 18830: gap of unknown length
 19376: contig of 546 bp in length
 19456: gap of unknown length
 20255: contig of 799 bp in length
 20335: gap of unknown length
 20577: contig of 622 bp in length
 21037: gap of unknown length
 22187: contig of 1150 bp in length
 22267: gap of unknown length
 22871: contig of 604 bp in length
 22951: gap of unknown length
 24044: contig of 1093 bp in length
 24124: gap of unknown length
 25189: contig of 1065 bp in length
 25269: gap of unknown length
 26217: contig of 948 bp in length
 26297: gap of unknown length
 27312: contig of 1015 bp in length
 28334: contig of 1542 bp in length
 29014: gap of unknown length
 29998: contig of 984 bp in length
 30078: gap of unknown length
 31456: contig of 1378 bp in length
 31536: gap of unknown length
 33242: contig of 1706 bp in length
 33322: gap of unknown length
 34819: contig of 1497 bp in length
 34899: gap of unknown length
 35765: contig of 866 bp in length
 35845: gap of unknown length
 37032: contig of 1187 bp in length
 37112: gap of unknown length
 38879: contig of 1767 bp in length
 38959: gap of unknown length
 39889: contig of 930 bp in length
 39969: gap of unknown length
 40967: contig of 998 bp in length
 41047: gap of unknown length
 42656: contig of 1609 bp in length
 42736: gap of unknown length
 44712: contig of 1436 bp in length
 44252: gap of unknown length
 45669: contig of 1417 bp in length
 45749: gap of unknown length
 47740: contig of 1991 bp in length
 47820: gap of unknown length
 48827: contig of 1007 bp in length
 48907: gap of unknown length
 49770: contig of 863 bp in length
 49850: gap of unknown length
 51951: contig of 2101 bp in length
 52031: gap of unknown length
 53654: contig of 1623 bp in length
 53734: gap of unknown length

```

* 53735 55071: contig of 1337 bp in length
* 55072 53151: gap of unknown length
* 56770: contig of 1619 bp in length
* 56850: gap of unknown length
* 56851 58743: contig of 1893 bp in length
* 58744 58823: gap of unknown length
* 58824 60291: contig of 1468 bp in length
* 60371: gap of unknown length
* 63759: contig of 3388 bp in length
* 63760 63839: gap of unknown length
* 63840 66476: contig of 2637 bp in length
* 66477 66556: gap of unknown length
* 66557 69322: contig of 2765 bp in length
* 69322 69401: gap of unknown length
* 69402 70806: contig of 1405 bp in length
* 70807 70886: gap of unknown length
* 70887 73742: contig of 2856 bp in length
* 73743 73822: gap of unknown length
* 73823 75073: contig of 1253 bp in length
* 75076 75155: gap of unknown length
* 75156 77979: contig of 2824 bp in length
* 77980 78059: gap of unknown length
* 81375 81376: contig of 3316 bp in length
* 81376 81455: gap of unknown length
* 81456 84432: contig of 2997 bp in length
* 84433 84532: gap of unknown length
* 84533 87596: contig of 3064 bp in length
* 87597 87676: gap of unknown length
* 87677 102529: contig of 14853 bp in length
* 102530 102609: gap of unknown length
* 102610 103128: contig of 519 bp in length
* 103129 103208: gap of unknown length
* 103209 103812: contig of 604 bp in length
* 103813 103892: gap of unknown length
* 103893 104112: contig of 220 bp in length
* 104113 104192: gap of unknown length
* 104193 104767: contig of 575 bp in length
* 104768 104847: gap of unknown length
* 104848 105570: contig of 723 bp in length
* 105571 105650: gap of unknown length
* 105651 106151: contig of 501 bp in length
* 106152 106231: gap of unknown length
* 106232 106941: contig of 710 bp in length
* 106942 107021: gap of unknown length
* 107022 107705: contig of 684 bp in length
* 107706 107785: gap of unknown length
* 107786 108412: contig of 627 bp in length
* 108413 108492: gap of unknown length
* 108493 109243: contig of 757 bp in length
* 109250 109329: gap of unknown length
* 109330 110040: contig of 711 bp in length
* 110041 110120: gap of unknown length
* 110121 110727: contig of 607 bp in length
* 110728 110807: gap of unknown length
* 110808 111413: contig of 606 bp in length

Query Match
Best Local Similarity 5.4%; Score 133.4; DB 29; Length 125448;
Matches 475; Conservative 0; Mismatches 436; Indels 30; Gaps 5;

Qy 1438 taccagaaatcatctctcttggaacagggctgccatcccgatgaagatcgaaatgtc 1497
Db 83158 TACCCGAAGATAATTTTCTGGCAGAGGCTCTGCAATCCCAATAAGACGCAACGCTG 83099

Qy 1498 agtccacactgtcaacataagcccgacacgtctctgactgagactggtgagggc 1557
Db 83098 AGCTCCATTCGATCAGACTGCAATTCATGCCCTACGGTGTGTTGGATTTGGAGAGGA 83039

Qy 1558 acattgggacgtgtgccttcattacgagaccag---gtggacagggctcctgtggacc 1614
Db 83038 ACTTATGTCACAAATTGACGACTATATGGACACGAGAAAGGCGATTCATTCCTCCGAC 82979

Qy 1615 ctggctgtgtttgtgtccacctgcaacgacgatcacccacacggcgttgccaaagtac 1674

```

```

Db 82978 CTACAGGCGATTTATGTATCCCAATTGTGATCCCGACCACCAATGGATTGATTGGTCTG 82919
Qy 1675 ttgtgcagagagaacgcgccttgcatctcttggaagacgccttcaccttctgtgtg 1734
Db 82918 CTCCGGGAAAG-----GAGGCAATTAAACCTAGGGCAGACCCACTTATTCTG 82871
Qy 1735 gttgcccccaaccagctcaaacgctggctccagcagtagtaccacaaccagtgccagaggtc 1794
Db 82870 CTGGCTCCTCGCAAAATTAACCTTGGTT---GGAGTTCTACAATCGACAAATAGAAACC 82814
Qy 1795 ctgcaccacatcagtatgatctcctgcccataatgccttcagggaagggctgagatctccagt 1854
Db 82813 GTAGAAGATGCGTATACTTCTAGTGGCAACGCTGAACCTAGCTAGTCCCTCGTACGCGGT 82754
Qy 1855 cctgcagtggaaagatgatcagttcgtctgttggaacatgattgattggaagagttcag 1914
Db 82753 GAACAAGTTGAACGCTAGGAATTACGTCCATATCCACTAGTCTAGTCTAGGACTGCCCC 82694
Qy 1915 acctgtctgtgctggcactgcaagcatgctttggctgtgcgtggtgcacacacctctggc 1974
Db 82693 AACTCTTTTCGGAATAAGCCTAACTCTGGCG-----GCAAAAGCAACAATAGCAACCC 82643
Qy 1975 tggaaagtgttctatttcoggggacacatgcctctgcgaggtctgtggtccgagtgggaaa 2034
Db 82642 GTCAAGATCACTTACAGCGCGGACACTATGCATGTCAAGATCTGATCGATCTGGCGCGC 82583
Qy 2035 gatccacctctctgatacatgaagcaccctggaagatggtttggaagagaagcaatg 2094
Db 82582 GACTCCACCGTCTTTATTCAGAGCGGACAATGGAGGATGACTTGGAGGAGGAGCGCGG 82523
Qy 2095 gaaagacacacagcacaaacgtcccaagccatcagctggggatgcggtgaaacgcgag 2154
Db 82522 TTAAGAGCGCACACACCGCTGTCGAGGCTATTCAGCAGGGCGGGAATATGAACCGCCGC 82463
Qy 2155 ttcattatgtgaaccacttcagcgcgtatgcctatgcctcctccctctt---cagcccc 2211
Db 82462 CATAAGATCTCTCCCACTTTTCGAGCGGTACGCTAAGTCCCGCGATACCGAGTGAT 82403
Qy 2212 aacttcagcagaaagtgaggagtgctcttgacacatgaaggtctgtctggagacttt 2271
Db 82402 GAGGATATGCAACGAGTTGCAATAGCCTTTGATATATGAAGTACCGCTTGAGGATCTG 82343
Qy 2272 ccaacaatgccccagctgattcccccaactgaaagccctgttctgtggtgcacatcgagag 2331
Db 82342 CAGCACTACCACACAGCTCTACCCCGCCCTTTTCGGATGTACGCGGAGTACACGGAAGAA 82283
Qy 2332 atgagagagcagggagagagcggagctgcggcaggtgcg 2372
Db 82282 CTGGAGCAGCGGCTGTCAAGCGGAGCTGAACAGGAGCG 82242

RESULT 6
AC007352 190574 bp DNA HTG 24-FEB-2000
LOCUS Drosophila melanogaster chromosome 2 clone BACR19J16 (D626) RPCI-98
DEFINITION 19,J16 map 47A-47B strain Y; cn bw sp, *** SEQUENCING IN PROGRESS
***, 36 unordered pieces.
AC007352
VERSION AC007352.8 GI:7025687
KEYWORDS HTG; HTGS_PHASE1.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 190574)
Celinker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hinkle,A., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K.,
Kearney,L., Lee,B., Lewis,S., Li,P., Ling,H., Moshrefi,A.R.,
Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B.,

```

Richards, S., Sethi, H., Svirskas, R.R., Wan, K.H., Webster, D.,
Woolley, P., Tang, S., Yee, M., Yu, C. and Rubin, G.M.
Sequencing of Drosophila melanogaster

TITLE
JOURNAL
REFERENCE
AUTHORS

2 (bases 1 to 190574)
Celniker, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazewicz, R.G.,
Butenoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,
Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,
Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L.,
Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P.,
Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S.,
Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Snir, E.,
Svirskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and
Rubin, G.M.

Direct Submission
Submitted (22-APR-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA

On Feb 24, 2000 this sequence version replaced qi:7018750.
For further information about this sequence replacement, including its location
and relationship to other sequences, please visit our sequence
archive web site (<http://www.fruitfly.org/sequence/>) or send email
to bdg@fruitfly.berkeley.edu. All contigs in this submission meet
the following cutoffs: length >= 200 bases.

* NOTE: This is a 'working draft' sequence. It currently
* consists of 36 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.

1 678: contig of 678 bp in length
* 679 758: gap of unknown length
* 759 1358: contig of 600 bp in length
* 1359 1438: gap of unknown length
* 1439 2373: contig of 935 bp in length
* 2374 2453: gap of unknown length
* 2454 2962: contig of 509 bp in length
* 2963 3042: gap of unknown length
* 3043 3889: contig of 847 bp in length
* 3890 3969: gap of unknown length
* 3970 5437: contig of 1468 bp in length
* 5438 5517: gap of unknown length
* 5518 6572: contig of 1055 bp in length
* 6573 6652: gap of unknown length
* 6653 8387: contig of 1735 bp in length
* 8388 8467: gap of unknown length
* 8468 9060: contig of 593 bp in length
* 9061 9140: gap of unknown length
* 9141 11167: contig of 2027 bp in length
* 11168 11247: gap of unknown length
* 11248 14518: contig of 3271 bp in length
* 14519 14598: gap of unknown length
* 14599 18837: contig of 4239 bp in length
* 18838 18917: gap of unknown length
* 18918 24324: contig of 5407 bp in length
* 24325 24404: gap of unknown length
* 24405 29984: contig of 5580 bp in length
* 29985 30064: gap of unknown length
* 30065 35794: contig of 5730 bp in length
* 35795 35874: gap of unknown length
* 35875 44808: contig of 8934 bp in length
* 44809 44888: gap of unknown length
* 44889 58222: contig of 13334 bp in length
* 58223 58302: gap of unknown length
* 58303 78322: contig of 20020 bp in length
* 78323 78402: gap of unknown length
* 78403 122249: contig of 43847 bp in length
* 122250 122329: gap of unknown length
* 122330 179102: contig of 56773 bp in length
* 179103 179182: gap of unknown length
* 179183 179407: contig of 225 bp in length
* 179408 179487: gap of unknown length

TITLE
JOURNAL

COMMENT

* 179488 180592: contig of 1105 bp in length
* 180593 180672: gap of unknown length
* 180673 181383: contig of 711 bp in length
* 181384 181463: gap of unknown length
* 181464 182092: contig of 629 bp in length
* 182093 182172: gap of unknown length
* 182173 183480: contig of 1308 bp in length
* 183481 183561: gap of unknown length
* 183562 183771: contig of 211 bp in length
* 183772 183851: gap of unknown length
* 183852 184058: contig of 207 bp in length
* 184059 184138: gap of unknown length
* 184139 185035: contig of 897 bp in length
* 185036 185116: gap of unknown length
* 185117 185846: contig of 731 bp in length
* 185847 185926: gap of unknown length
* 185927 186560: contig of 634 bp in length
* 186561 186640: gap of unknown length
* 186641 187533: contig of 613 bp in length
* 187534 187611: gap of unknown length
* 187612 187691: gap of unknown length
* 187692 188409: contig of 718 bp in length
* 188410 188489: gap of unknown length
* 188490 189196: contig of 707 bp in length
* 189197 189276: gap of unknown length
* 189277 189969: contig of 693 bp in length
* 189970 190049: gap of unknown length
* 190050 190574: contig of 525 bp in length.

FEATURES
Location/Qualifiers

1..190574
/organism="Drosophila melanogaster"
/strain="y; cn bw sp"
/db_xref="taxon:7227"
/chromosome="2"
/map="47A-47B"
/clone="BACR19J16 (D626) RPCI-98 19.J.16"
/clone_lib="RPCI-98 (Roswell Park Cancer Institute
Drosophila melanogaster BAC library, partial EcoRI in
pBACs3.6"

BASE COUNT 53370 a 39726 c 40358 g 54299 t 2821 others
ORIGIN

Query Match 5.4%; Score 133.4; DB 29; Length 190574;
Best Local Similarity 50.5%; Pred. No. 6.2e-20;
Matches 475; Conservative 0; Mismatches 436; Indels 30; Gaps 5;

QY 1438 taaccagaaatcatcttccttggaacagggtctgccatcccgatgaagattcgaatgtc 1497
DB 132812 TACCCGAAGATAATTTCTTGGGCACAGGCTCTGCTATCCCAATAAGACGCGAAACGTG 132871
QY 1498 agtgcacacattgtcaacataaagccccgcacagctctctactgactgtgtgagggc 1557
DB 132872 AGCTCCATTCTGATCAGGACTGCAATTGATGCTTGTGTTGGATTGTGAGAGGA 132931
QY 1558 acattgggcagctgtgcctgcatctacggagaccag---gtgcacaggggtccctgggcacc 1614
DB 132932 ACTTATGGTCAAAATTTGTACGACTATATGGACACACCAAGGCGCATTTGTTCCCGAC 132991
QY 1615 ctggctgctgtgttctccacacctgcacgcagatcacacacagggcttgcgaagtac 1674
DB 132992 CTACAGGCGATTTATGTATCCCATTTGATCGCGACACACACATTTGATTTGGTCTG 133051
QY 1675 ttgctgcagagagaaacgccttggcatcttttgggaaagccgtctcaccccttgcctggtg 1734
DB 133052 CTCCTGGGAAAG-----GAGGCAATTAACCTAGGCGACAGCCCATTTATCTG 133099
QY 1735 gttgcccccaaccagctcaaacgtctgcctccagcagatcacacacagcagtgccaggggtc 1794
DB 133100 CTGGCTCTCGCCAAATTAACCTTGGTT---GGAGTCTTACAAATGACAAATAGAAACC 133156
QY 1795 ctgcaccacatcagatgattctctgccttgccttcaggagggggtgagatctccagt 1854

after the database hits. Genes without significant peptide similarity but with ESI similarity are named as 'unknown' proteins. Genes without protein or EST similarity, that are predicted by more than two gene prediction programs over most of their length are annotated as 'hypothetical' proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats are identified by repeatmasker (Arian Smit, <http://ftp.genome.washington.edu/RM/RepeatMasker.html>).

FEATURES

source	1. .46335 /organism="Arabidopsis thaliana" /cultivar="Columbia" /db_xref="taxon:3702" /chromosome="1" /map="m213" /clone="F915"	repeat_region	13176. .13409 /note="novel repeat from T25M19 22823 to 23610 788 nt"
misc_feature	1. .5778 /note="overlap with BAC clone F5F19 (AC008216:105116. .110893)."	repeat_region	13177. .13298 /note="novel repeat from T25M19 22823 to 23610 788 nt"
misc_feature	1 (AC008216:105116. .110893)."	repeat_region	15264. .15339 /rpt_family="(GAAA)n" complement(18843. .18918) /rpt_family="(GAAA)n"
repeat_region	complement(1497. .1591) /rpt_family="(TAAAA)n"	repeat_region	19668. .19760 /rpt_family="POLY_A" complement(20890. .20985) /rpt_family="(GA)n"
repeat_region	complement(2125. .2194) /rpt_family="(GAAA)n"	repeat_region	complement(20931. .20985) /rpt_family="(GAA)n"
mRNA	9063. .9373,9467. .9826,9904. .9947,10041. .10095,10177. .10221,10306. .10395,10478. .>10639)) /gene="F915.1" complement(<7284. .>10639) /gene="F915.1"	mRNA	join(<21044. .21140,21233. .21450,23180. .23739,24098. .25199) /gene="F915.4"
gene	complement(<7284. .>10639) /gene="F915.1"	gene	<21044. .25199 /gene="F915.4"
CDS	complement(join(7284. .7610,7681. .8652,8869. .9016,9063. .9373,9467. .9826,9904. .9947,10041. .10095,10177. .10221,10306. .10395,10478. .>10639)) /gene="F915.1" /note="hypothetical protein"	CDS	join(21044. .21140,21233. .21450,23180. .23739,24098. .25046) /gene="F915.4"
repeat_region	complement(<7284. .>10639) /product="putative peptide transporter"	repeat_region	/codon_start=1 /product="putative peptide transporter"
repeat_region	complement(2125. .2194) /rpt_family="(GAAA)n"	repeat_region	/protein_id="AAF29404.1" /db_xref="GI:6850341"
mRNA	9063. .9373,9467. .9826,9904. .9947,10041. .10095,10177. .10221,10306. .10395,10478. .>10639)) /gene="F915.1" /note="hypothetical protein"	mRNA	26156. .26179 /rpt_family="(TA)n"
gene	complement(<7284. .>10639) /gene="F915.1"	gene	26779. .26899 /rpt_family="(CAAAA)n"
CDS	complement(join(7284. .7610,7681. .8652,8869. .9016,9063. .9373,9467. .9826,9904. .9947,10041. .10095,10177. .10221,10306. .10395,10478. .>10639)) /gene="F915.1" /note="hypothetical protein"	CDS	join(<26899. .26982,27574. .>27987) /gene="F915.5"
repeat_region	complement(<7284. .>10639) /product="putative peptide transporter"	repeat_region	<26899. .>27987 /gene="F915.5"
repeat_region	complement(2125. .2194) /rpt_family="(GAAA)n"	repeat_region	/note="predicted by genscan"
mRNA	9063. .9373,9467. .9826,9904. .9947,10041. .10095,10177. .10221,10306. .10395,10478. .>10639)) /gene="F915.1" /note="hypothetical protein"	mRNA	join(26899. .26982,27574. .27987) /gene="F915.5"
gene	complement(<7284. .>10639) /gene="F915.1"	gene	/codon_start=1 /protein_id="AAF29405.1"
CDS	complement(join(7284. .7610,7681. .8652,8869. .9016,9063. .9373,9467. .9826,9904. .9947,10041. .10095,10177. .10221,10306. .10395,10478. .>10639)) /gene="F915.1" /note="hypothetical protein"	CDS	/db_xref="GI:6850342"
repeat_region	complement(<7284. .>10639) /product="putative peptide transporter"	repeat_region	/translation="MMRYCOFTPYDPSSNMDDLKAFEDETWKYKKKKAWSLFEVLY SEEMKNMGEYTLNIMVNECFKIGRFQDAIQIFYTKALKLKNPDVTCYRNITTLNKG GVLFEAFYFEEMCSDRLLVPPDVFTYTTMIDAYLKAGKTEDALRISKNMWDALFGVA WLACL"
repeat_region	complement(2125. .2194) /rpt_family="(GAAA)n"	repeat_region	complement(27450. .27491) /rpt_family="(TA)n"
mRNA	9063. .9373,9467. .9826,9904. .9947,10041. .10095,10177. .10221,10306. .10395,10478. .>10639)) /gene="F915.1"	mRNA	complement(join(<29019. .29311,29402. .29473,29582. .>29780)) /gene="F915.6"
gene	complement(<7284. .>10639) /gene="F915.1"	gene	complement(<29019. .>29780) /gene="F915.6"
CDS	complement(join(7284. .7610,7681. .8652,8869. .9016,9063. .9373,9467. .9826,9904. .9947,10041. .10095,10177. .10221,10306. .10395,10478. .>10639)) /gene="F915.1" /note="hypothetical protein"	CDS	/note="similar to PGP224 GB:AAD02554 from [Petunia x hybrida]" complement(join(29019. .29311,29402. .29473,29582. .29780))

Protein coding regions (CDS) have been predicted with the help of computer analysis using the Genefinder program in PomBase (an ACEDB database) with additional predictions for the branch-acceptor sites supplied by the program Splice. CAUTION: It is possible that for any individual CDS we may have underestimated or overestimated the number of introns/exons or we may not have chosen the correct splice donor/acceptor sites.

CDS are numbered using the following system eg SPBC25H2.01c. SP (S. pombe), B (chromosome 2), c25H2 (cosmid name), .01 (first CDS), c (complementary strand).

The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous.

The length in codons is given for each CDS.

IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions.

FEATURES

location/Qualifiers

source

1. .32329

/organism="Schizosaccharomyces pombe"

/strain="972h"

/db_xref="taxon:4896"

/chromosome="II"

/clone="cosmid c3D6"

/map="IIR"

gene

1. .359

/gene="SPBC3D5.01"

CDS

1. .359

/partial

/gene="SPBC3D5.01"

/note="SPBC3D5.01, len:118aa"

/codon_start=3

/label="SPBC3D5.01"

/product="hypothetical protein"

/protein_id="CAB09128.1"

/db_xref="GI:2117297"

/db_xref="SWISS-PROT:P87166"

/translation="SFPYHFSFLSPILVFLPHSFALPSTSISSRFLASPHQSLLFS
FAARIHPSALHVLSEVRFAYHRCPRSVALTTPACRFVPPSSDVNHAIGRTSGREVF
NSLPSPVHGSDTEL"

gene

1655. .287

/gene="SPBC3D6.02"

CDS

1655. .287

/gene="SPBC3D6.02"

/note="SPBC3D6.02, len:390, SIMILARITY:Schizosaccharomyces
pombe, 042663, hypothetical 39.7 kd protein., (383 aa),
fast a scores: opt: 301, E():9.8e-13, (22.7% identity in
396 aa)"

/codon_start=1

/label="SPBC3D6.02"

/product="hypothetical protein."

/protein_id="CAB09122.1"

/db_xref="GI:2117298"

/db_xref="SWISS-PROT:P87167"

/translation="MQLNSFLGFAASTAVLASSADAAPTLYKRKSKSNFTAKSAVAI
LDGVNNTFGVISTREKPPMDYHIWHTASTPGTKYLPWNDVAPSTYIDEGFTLR
TGKRAYLGDNMDLAFTNHTDKASHDDGTHRVLSQRKANFMATQRCGLDPFG
YOLKRSKVFQAGQVFGSGRSIDCWMDSVLNLGRYGNTEALSPLPRNLISA
DIPAKSRLFLPHGYINFNFSAPNKRMRFTASEATTVNGQNSVLTVDVPRSLIY
YTSCALERFTNEFFPDVTPNDGSAQFVVYNMGNPKQTITSSKGPRLYEVARPNC
YTRGCEYTNICPRAHSHYELAPASFNISLWISQISPRIGVTLQVYINANED"

/complement(3885..5921)

/gene="SPBC3D6.03c"

gene

complement(3885..5921)

CDS

/gene="SPBC3D6.03c"

/note="SPBC3D6.03c, len:678,

SIMILARITY:Schizosaccharomyces pombe, YATA_SCHPO,

hypothetical 90.6 kd protein cid4.10 in chromosome i.,

(809 aa), fasta scores: opt: 280, E():1.8e-10, (32.0%

identity in 494 aa)"

/codon_start=1

/label="SPBC3D6.03c"

/product="conserved hypothetical protein"

/protein_id="CAB09123.1"
/db_xref="GI:2117299"
/db_xref="SPTREMBL:P87168"
/translation="MKASLLVPRALLFGQLPPKYSWYSVKRWOSQLTFRNKRKNT
NRIMSVSSLLNPDSLIAPLLCVDLDNRKGLJSGMGLTQMKFRSQAGVSGKSVSVF
LMPPLSLNANGTITAGLFGYLOSSGIONTWGLHAPKPVISIIKKSHHLLFSGPLRLD
LNSFSEONADATNSFYLDPEFTIKGNIYNSWSFLSPNSKEAGVFNADKALALGV
PFGNGKAGEAVLSKDGTTWYIYQVGPGRKQYFYVLGCSLSALNQMSKHYD
SPSDVPTCIHILEKIGWPEYIKFLSHPKFSRAQHFISCIELASNNPFGNKGKN
VLACRDFAADIKPSTLDTQLPENTVLAKEETSMVLYDECKISESPSYSPVKLA
KFSFENLPPFENEGETYLDVLTGTSATCPWRBSLSYSVAIDGTVMIDCGEGATSQF
FRQYTNTEPMLRKLKAIFITHLSDHYLGLNLVQWNRKANTNNSMHIINIGPKFLW
QWLQRLKSPANLOALLNRIIFIATKETTPTLQLTSDSISISSVTHINDSCLISH
TRYKGLVSGDTRPNKELVKAGISLGLHSTFEDDLKHEAIQHQHSTASALSYAQ
SNKAKALILTHFSQSYDAFLPPDWTIYPKSKTIYANDGLOWQOQFSQKRETI"
complement(4239..4268)
/gene="SPBC3D6.03c"
/note="PS00152 ATP synthase alpha and beta subunits
signature"
complement(5628..5651)
/gene="SPBC3D6.03c"
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"
join(6157..6159,6210..6450,6555..6577)
/gene="rps25-1"
/note="SPBC3D6.15"
join(6157..6159,6210..6450,6555..6577)
/gene="rps25-1"
/note="SPBC3D6.15, len:92"
/codon_start=1
/label="rps25"
/product="40s ribosomal protein s25"
/protein_id="CAB09129.2"
/db_xref="GI:4581499"
/translation="MAPKKKSKGKVKDKAQHATVFDKSIIDRINKEVPAKFISVS
LVDRMKINGSLARIAIRDLAERGVIQKVQDHSKQALYTRAAASA"
/6160..6165
/note="gtaagt, splice donor sequence"
/6192..6209
/note="ctaattccattattag, splice branch and acceptor
sequence"
6451..6554
/note="confirmed"
complement(join(6886..8804,8855..9005))
/gene="SPBC3D6.04c"
complement(join(6886..8804,8855..9005))
/note="SPBC3D6.04c"
/note="SPBC3D6.04c, len:689, SIMILARITY:Saccharomyces
cerevisiae, MAD1_YEAST, spindle assembly checkpoint
component ma d1, (749 aa), fasta scores: opt: 454,
E():1.1e-16, (23.4% identity in 710 aa)"
/codon_start=1
/label="SPBC3D6.04c"
/product="hypothetical protein"
/protein_id="CAB09124.1"
/db_xref="GI:2117300"
/db_xref="SPTREMBL:P87169"
/translation="MSSKLTIVQATTSMADSPRDFQSRSQLPRFLATSVKKPNLKKP
SVNSANETNPKLASLEFOLNKLNDLKRKEFEFEQTELORKLAEHEQKNSLQLR
LTLVEKQLEQOSTSYQKLEEEVNEKEATQVKIHLLDAKWEIHELTKQIEKNDQAL
SKKNHEVSWAOLKOTNLNLEKFLADSRQLETRCKLEAAAEQQLSELVHNQO
LEESTQVSSSIELEKINAEQLQISELEKLSKAAQEEERLEKLSNNRNVEILKEKND
LESKLYFEYRDVATLELENERIOTELNSWKSLITNELPTPEAVSNKLVFLQNTNA
NLGERVSSLESQLSNKPANPLGANEKAAHITETELTKELHEQNRRRLQROKSLATQ
EIDLRLNLSYDDREALISEKNTMKKLETEGLVKLVDEYKLVKLESPVSLDVEDT
SDEVSLOKRRKNEHKGADGYTLYKKNQHLIFOVKEKTNIEAFIREQILTLESSLAT
LRQELAQVTEINSCVLRHSNPTIKYERIKAAQLEMLNAENSALKALEDKKVOCLP
IQSFIAERKALDKLKEVAEREKRIEFSVKSLEFEAVFLGKGLDFMPNGS
VAVTYSYREDNTAFIFDGSSTMKLVGNPSPGPEFLIRFWCDERKKTIPGMLAALT
ELLDKND"
complement(8805..8818)
/note="ctgacaccttag, splice branch and acceptor
sequence"
complement(8849..8854)

misc_feature

misc_feature

gene

CDS

misc_feature

misc_feature

intron

gene

CDS

misc_feature

misc_feature

RESULT 15

AE001039 23533 bp DNA BCT 15-DEC-1997
LOCUS Archaeoglobus fulgidus section 68 of 172 of the complete genome.
DEFINITION AE001039 AE000782
ACCESSION AE001039.1 GI:2689362
KEYWORDS

ORGANISM Archaeoglobus fulgidus.
SOURCE Archaeoglobus fulgidus
ARCHAEA; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;

Archaeoglobus.
1 (bases 1 to 23533)

REFERENCE AUTHORS Klenk, H.P., Clayton, R.A., Tomb, J., White, O., Nelson, K.E.,

Ketchum, K.A., Dodson, R.J., Gwinn, M., Hickey, E.K., Peterson, J.D.,

Richardson, D.L., Kierulff, A.R., Graham, D.E., Kyrpides, N.C.,

Fleischmann, R.D., Quackenbush, J., Lee, N.H., Sutton, G.G., Gill, S.,

Kirkness, E.F., Dougherty, B.A., McKenney, K., Adams, M.D., Loftus, B.,

Peterson, S., Reich, C.I., McNeill, L.K., Badger, J.F., Glodek, A.,

Zhou, L., Overbeek, R., Gocayne, J.D., Weidman, J.F., McDonald, L.,

Utterback, T., Cotton, M.D., Spriggs, T., Artach, P., Kaine, B.P.,

Sykes, S.M., Sadow, P.W., D'Andrea, K.P., Bowman, C., Fujii, C.,

Garland, S.A., Mason, T.M., Olsen, G.J., Fraser, C.M., Smith, H.O.,

Woese, C.R. and Venter, J.C.

The complete genome sequence of the hyperthermophilic,
sulphate-reducing archaeon Archaeoglobus fulgidus

Nature 390 (6558), 364-370 (1997)

98049343
2 (bases 1 to 23533)

REFERENCE AUTHORS Klenk, H.P., Clayton, R.A., Tomb, J., White, O., Nelson, K.E.,

Ketchum, K.A., Dodson, R.J., Gwinn, M., Hickey, E.K., Peterson, J.D.,

Richardson, D.L., Kierulff, A.R., Graham, D.E., Kyrpides, N.C.,

Fleischmann, R.D., Quackenbush, J., Lee, N.H., Sutton, G.G., Gill, S.,

Kirkness, E.F., Dougherty, B.A., McKenney, K., Adams, M.D., Loftus, B.,

Peterson, S., Reich, C.I., McNeill, L.K., Badger, J.F., Glodek, A.,

Zhou, L., Overbeek, R., Gocayne, J.D., Weidman, J.F., McDonald, L.,

Utterback, T., Cotton, M.D., Spriggs, T., Artach, P., Kaine, B.P.,

Sykes, S.M., Sadow, P.W., D'Andrea, K.P., Bowman, C., Fujii, C.,

Garland, S.A., Mason, T.M., Olsen, G.J., Fraser, C.M., Smith, H.O.,

Woese, C.R. and Venter, J.C.

Direct Submission
Submitted (15-DEC-1997) The Institute for Genomic Research, 9712

Medical Center Dr, Rockville, MD 20850, USA

In order to show the genes in ascending order on the genome, the

origin of this version has been moved by TIGR to position 2093570

of the original version and the opposite strand is shown from the

original version.

On Dec 16, 1997 this sequence version replaced gi:2649651.

COMMENT Location/Qualifiers
FEATURES

1..23533
/organism="Archaeoglobus fulgidus"
/db_xref="taxon:2234"
complement(125..757)
/gene="AF0928"
complement(125..757)
/gene="AF0928"
/note="similar to PID:1050822 PID:1164975 PID:1420338
percent identity: 27.60; identified by sequence
similarity: putative"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AAB90310.1"
/db_xref="GI:2649668"
/translation="MHQVIOQRKPESSGESSRFELIAFDLDTLIPFDEIRREL
GIEGRFLESIMDEKDLQROEMLRVLESRCAMEARPAYAYAVELVNGLDATKGV
TNSRSEIIVAEIRLGFEDFLIGREDAPKPSPEPLNLRMFEDVSPSKALMGVDFL
FDLUSGKAAGAKTALITLDRNGMAKSFQIADYVFSKLADFLFEVEQ"
complement(699..1424)
/gene="AF0929"
complement(699..1424)
/gene="AF0929"
/note="similar to GB:L77117 SP:Q57700 PID:1590982 percent
identity: 46.67; identified by sequence similarity;

putative"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AAB90309.1"
/db_xref="GI:2649667"
/translation="MVVONGIKLLRKALKINAGEKVIPTVQOQTMKQLIALDVM
GEKAMEIAKVAEHVDRIKVNPLVLSAGVIMKRLSEIKPVIADFKIADVPYTSLLI
ARIAPENAESVIVHGFYGSOTLREVCRVAEEFGKVVAVTELSPPGGEEMSAVSLK
IVERAKAEGCHLLAPSTRIRLRIRKAAGDMETLCPGICGAQKGSIEAVKYADGIIV
GRGIYASGNPAEARKLRVLKI"
1430..2674
/gene="AF0930"
1430..2674
/gene="AF0930"
/note="similar to GB:L77117 SP:Q58080 PID:1591380 percent
identity: 44.81; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=1
/product="molybdenum cofactor biosynthesis protein
(moaA-2)"
/protein_id="AAB90308.1"
/db_xref="GI:2649666"
/translation="MVREKGFWRERDEALYRLKKIKALETEDSVSVNLQGRVLAE
DIVAGDFVPEFNRAAMDGYAKAEDTFGASVSNPIMQLAGSVEIGEVPKVVESGMA
VRIMTGMMPCTNAVMLEHTRLNGNEVEVLKSVPMKNVSRDEGAAAGVGVFRKG
ETLPQDAGVIAISLGIESRVVRKPRVAVITGNELAEVGEPLGAKIYNSNPICN
ALLERGFPAISIAIDDEELRNKLOKALFEADAVITGTSVGAADLVDPVVGEGE
IVFVGVMKPMGPGTAAVVEGKPMPLPGSPAAALLGFTYFAIPALYRMVNRITARK
WSRGVQLQGRIPSEIGVRNVLWEDGKVPIRISGSLSEFVRANALLVVPEDK
EGYEGEVEVETLLRDVTEVE"
2671..4635
/gene="AF0931"
2671..4635
/gene="AF0931"
/note="similar to GB:L77117 SP:Q58296 PID:1591564 percent
identity: 50.83; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=1
/product="molybdenum cofactor biosynthesis protein
(moaA-1)"
/protein_id="AAB90307.1"
/db_xref="GI:2649665"
/translation="MRKVFQDVTVDEARGLLFNFSPRRKTEVVPLHLAAGRVLAKD
VYALTDVPFORATMDGYAVRAEDTFEAEEDSPVTLKVGTVEACKRPELEVESKAV
EIAATGMMPKANAVVMVETFRREDEIDFKVPSPGNVMTAGSDVWAGELIAREGT
ELTHREIGVMAACGISEVEYCPKPAVITSTNELVKPGEIGEKIYDVNSYIAAA
VEENGGEAILMGIAIDNEEMRSIIAKALREADIVITSGTSAGADVMYRILDDFHP
GVIVHGAIKPKCPAVIAVSDSKPIFGPLPYPTSAMTVFEVIVAPIIRELAGRESSG
AKVRARLAIKVFSAEGRREYLPVNVVEGAGYSAYPVSGSYSCAVTAFATDGFVEIP
ENVVLMRGEDEVEVKLFSLKPADLMIITGSHICIGVDILLSMRKKKPYTSKVVNGST
GGILAVKRGADIAHTLDESQVYNEPIRVRYGVDRVAVLVKGYLREGILVAKGNPK
SITGFEDLLREDVTFINRNSGTRILDMYLRVAREKRGSLSFNELTSIKGYSIEAK
THTSVAVASGADGVGIRSVASQYGLDFILRSEYDFILIRDRLEKKAVIDFLE
CLTSEFAKELERVVGLRVYERTREVIED"
4677..6677
/gene="AF0932"
4677..6677
/gene="AF0932"
/note="similar to GB:L77117 SP:Q58010 PID:1591298 percent
identity: 34.43; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AAB90306.1"
/db_xref="GI:2649664"
/translation="MRLLLEHESKALLEKYGITAKCIFCETEEAOVKAKEIGPPVV
MKVAGREILHKSDVGVGVILNVKSEDEVEVFORLMSIPKAEGVNLQPOLEKGVIEIG
VAENPQFGSVAMFGLGVGVFEVLKDVFSRLLPLTRDAEEMVRVKGKYLEGIVGK
GDVGAVDILLKLNVEVEKEGIVEMDLNPNFVYERGAVVVDARIYVGERKRFEMDVG

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 18, 2001, 05:20:16 ; Search time 128.59 Seconds
(without alignments)
7247.991 Million cell updates/sec

Title: US-09-434-382-1
Perfect score: 2481
Sequence: 1 atcggcgcttgcgcgt.....agaaggtcagagccagtgta 2481

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 480022 seqs, 187831343 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_36: *
1: /cgnl_8/gcgdata/geneseq/geneseqn/NA1980.DAT: *
2: /cgnl_8/gcgdata/geneseq/geneseqn/NA1981.DAT: *
3: /cgnl_8/gcgdata/geneseq/geneseqn/NA1982.DAT: *
4: /cgnl_8/gcgdata/geneseq/geneseqn/NA1983.DAT: *
5: /cgnl_8/gcgdata/geneseq/geneseqn/NA1984.DAT: *
6: /cgnl_8/gcgdata/geneseq/geneseqn/NA1985.DAT: *
7: /cgnl_8/gcgdata/geneseq/geneseqn/NA1986.DAT: *
8: /cgnl_8/gcgdata/geneseq/geneseqn/NA1987.DAT: *
9: /cgnl_8/gcgdata/geneseq/geneseqn/NA1988.DAT: *
10: /cgnl_8/gcgdata/geneseq/geneseqn/NA1989.DAT: *
11: /cgnl_8/gcgdata/geneseq/geneseqn/NA1990.DAT: *
12: /cgnl_8/gcgdata/geneseq/geneseqn/NA1991.DAT: *
13: /cgnl_8/gcgdata/geneseq/geneseqn/NA1992.DAT: *
14: /cgnl_8/gcgdata/geneseq/geneseqn/NA1993.DAT: *
15: /cgnl_8/gcgdata/geneseq/geneseqn/NA1994.DAT: *
16: /cgnl_8/gcgdata/geneseq/geneseqn/NA1995.DAT: *
17: /cgnl_8/gcgdata/geneseq/geneseqn/NA1996.DAT: *
18: /cgnl_8/gcgdata/geneseq/geneseqn/NA1997.DAT: *
19: /cgnl_8/gcgdata/geneseq/geneseqn/NA1998.DAT: *
20: /cgnl_8/gcgdata/geneseq/geneseqn/NA1999.DAT: *
21: /cgnl_8/gcgdata/geneseq/geneseqn/NA2000.DAT: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	237	9.6	238	21	280231
2	51.4	2.1	1501	21	AI2583
3	48.2	1.9	799	19	V55831
4	48.2	1.9	1925	20	X90924
5	48.2	1.9	5452	20	X90923
6	48.2	1.9	8705	20	Z23778
7	48.2	1.9	9600	19	V21683
8	48.2	1.9	10380	20	Z22248
9	48.2	1.9	10596	14	Q51731
10	48.2	1.9	10596	17	T40348
11	48.2	1.9	10596	20	X15650
12	45	1.8	795	19	V55830

13	43.8	1.8	1528	20	X35555	Secreted protein c
14	43.6	1.8	217	16	Q97711	Branch-chain alp
15	43.4	1.7	1567	19	X03145	Aspergillus terreu
16	43	1.7	2125	20	X36279	Wheat Rht clone 5a
17	42.8	1.7	453	20	X36263	DNA sequence obtai
18	42.8	1.7	2709	20	X36275	Composite DNA sequ
19	42.8	1.7	16442	18	X83006	Partial mouse WRN
20	42.6	1.7	10732	21	AI0594	Gene encoding a su
21	42.4	1.7	324	20	X36273	DNA sequence obtai
22	42	1.7	2297	21	X93309	Sequence encoding
23	41.4	1.7	38186	20	Z32028	Human MTH1 relate
24	41	1.7	2301	19	V20445	Human c-trk oncoge
25	41	1.7	2301	21	A09303	Human Trk oncogene
26	40.8	1.6	1429	21	Z93300	Sequence encoding
27	40.4	1.6	15377	13	Q25975	MH mutant porcine
28	40.2	1.6	1412	17	T26989	Thermus aquaticus
29	39.8	1.6	6306	15	Q54841	Sequence encoding
30	39.8	1.6	6579	18	T77783	cDNA encoding LexA
31	39.8	1.6	6819	18	T77782	cDNA encoding GAL4
32	39.6	1.6	730	19	V62161	HSV-2 strain SB5 C
33	39.6	1.6	925	21	Z52534	Human secreted pro
34	39.6	1.6	2744	16	Q98470	Mispl1-containing p
35	39.6	1.6	4289	19	V62147	HSV-2 strain SB5 C
36	39.6	1.6	9551	20	Z22301	cDNA encoding a hu
37	39.6	1.6	51259	18	X83007	Partial mouse WRN
38	39.6	1.6	117213	19	V62176	HSV-2 strain SB5 C
39	39.4	1.6	1675	20	Z32195	Corn hexose carrie
40	39.4	1.6	4020	18	T91361	Orf virus genomic
41	39.4	1.6	114955	20	X53491	Human adenosine A1
42	39	1.6	1994	21	A35027	Human adenosine re
43	39	1.6	2009	20	Z07513	Human RIP-associat
44	39	1.6	2034	20	Z07514	Human RIP-associat
45	39	1.6	8631	21	A35028	Human adenosine re

ALIGNMENTS

RESULT	1
Z80231	
ID	Z80231 standard; cDNA; 238 BP.
XX	Z80231;
AC	Z80231;
DT	07-APR-2000 (first entry)
XX	Human colon cancer cell line SW480 cDNA clone SEQ ID NO:315.
DE	Human; gene expression product; diagnosis; tumour; colon cancer;
XX	colorectal adenocarcinoma; cell line SW480; cell proliferation;
KW	cytostatic; sarcoma; breast cancer; neoplasia; dysplasia;
KW	hyperplasia; ds.
XX	Homo sapiens.
OS	XX
PN	WO9964576-A2.
XX	XX
PD	16-DEC-1999.
XX	XX
PF	09-JUN-1999; 99WO-IB01062.
XX	XX
PR	10-JUN-1998; 98US-0088801.
XX	(FARB) BAYER CORP.
PA	Endege WO, Steinmann KE, Astle JH, Burgess CC, Bushnell SE;
PI	Carroll E, Catino TU, Derti A, Ford DM, Lewis ME, Monahan JE;
PI	Schlegel R;
XX	WPI; 2000-087220/07.
DR	Novel nucleic acids, used to develop products for the diagnosis and
XX	treatment of disorders involving unwanted cell proliferation,
PT	

PT	particularly cancers, especially colon cancer	-
XX	Claim 15; Page 258; 469pp; English.	
PS		
XX		
CC	279917 to 280766 represent double stranded cDNA clones isolated from the	
CC	human colorectal adenocarcinoma (colon cancer) cell line SW480. The	
CC	cDNA clones can be used to generate antisense oligonucleotides which	
CC	can be used for antisense therapy. Methods and products from the present	
CC	invention can be used for identifying and/or classifying cancerous cells	
CC	present in a human tumour, particularly in solid tumours, e.g. carcinomas	
CC	and sarcomas, e.g. breast or colon cancers. The cDNA clones can be used	
CC	for developing agents for the diagnosis and treatment of disorders	
CC	involving unwanted cell proliferation, such as neoplasia, dysplasia or	
CC	hyperplasia.	
XX	Sequence 238 BP; 55 A; 57 C; 69 G; 57 T; 0 other;	
SQ		

Query Match	9.6%	Score 237	DB 21	Length 238
Best Local Similarity	100.0%	Prod. No. 2.6e-51		
Matches 237	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY	176	acctgcaggtggtgagcgggtagccggagctcggcgccgagctctagctctctccg	235	
Db	1	acctgcaggtggtgagcgggtagccggagctcggcgccgagctctagctctccg	60	
QY	236	agttcaaccggatctcttcaactgtggagaagcggttcagagactcatcaggagcacca	295	
Db	61	agttcaaccggatctcttcaactgtggagaagcggttcagagactcatcaggagcacca	120	
QY	296	agttaaaggtgtcgcctcgagcaacatactctgcacacaaatgcactgtctaattgtg	355	
Db	121	agttaaaggtgtcgcctcgagcaacatactctgcacacaaatgcactgtctaattgtg	180	
QY	356	ggggcttaagtgaatgatctctactttaaggaaaccggcttccaaagtgtgtac	412	
Db	181	ggagctcaactgaatgatctctactttaaggaaaccggcttccaaagtgtgtac	237	

RESULT	2	
AL12583	AL12583 standard; cDNA; 1501 BP.	
XX	XX	
XX	AC	
XX	AC	
XX	XX	
XX	XX	
XX	25-JUL-2000 (first entry)	
XX	XX	
DE	cDNA encoding a membrane associated organizational protein (HJNCT).	
XX	XX	
XX	Human; membrane associated organizational protein; HJNCT;	
KW	cell proliferative disorder; cancer; autoimmune disorder;	
KW	inflammatory disorder; neurological disorder; developmental disorder;	
KW	vesicle trafficking; reproductive disorder; gastrointestinal disorder;	
KW	renal disorder; atherosclerosis; leukaemia; rheumatoid arthritis;	
KW	Alzheimer's disease; anxiety; diabetes; ovulatory defect; renal failure;	
KW	irritable bowel syndrome; allergy; ss.	
XX	XX	
XX	Homo sapiens.	
XX	XX	
Key	Location/Qualifiers	
FH	183..1268	
FT	FT	
FT	/*tag= a	
FT	/product= "membrane associated organizational protein"	
XX	XX	
XX	WO200018915-A2.	
XX	XX	
XX	06-APR-2000.	
XX	XX	
XX	23-SEP-1999;	99WO-US22082.
XX	XX	
XX	25-SEP-1998;	98US-0155215.
XX	13-OCT-1998;	98US-0155251.
XX	04-MAY-1999;	99US-0172228.
XX	XX	

(INCY-) INCYTE PHARM INC.

Yue H, Lal P, Corley NC, Guegler KJ, Baughn MR, Lu AD, Tang YT;
WPI; 2000-293154/25.
P-PSDB; Y84607.

Human membrane associated organizational protein and nucleic acid
sequences useful in the diagnosis, treatment and prevention of cell
proliferative associated disorders e.g. cancer, rheumatoid arthritis
and Alzheimer's disease -

Claim 9; Page 78; 84pp; English.

The present sequence encodes a human membrane associated organizational
protein (HJNCT). HJNCT is used for the diagnosis, treatment and
prevention of cell proliferative disorders including cancer and
autoimmune/inflammatory, neurological, developmental, vesicle
trafficking, reproductive, gastrointestinal and renal disorders. These
disorders may include atherosclerosis, leukaemia, allergies, rheumatoid
arthritis, Alzheimer's disease, anxiety, diabetes, ovulatory defects,
renal failure and irritable bowel syndrome. A vector expressing HJNCT,
and an agonist of HJNCT can be used to treat or prevent a disorder
associated with decreased expression or activity of HJNCT. An antagonist
of HJNCT or a vector expressing the complement of a polynucleotide
encoding HJNCT can be used to treat or prevent a disorder associated
with increased expression or activity of HJNCT. Antibodies which bind
HJNCT can be used for diagnosis of disorders associated with HJNCT
expression or to monitor patients being treated with HJNCT, agonists,
antagonists or inhibitors of HJNCT. Assays are preferably carried out on
body fluids from a patient using radioimmunoassay, enzyme linked
immunosorbent assays or fluorescent activated cell sorting assays.
Polynucleotides encoding HJNCT are also used in hybridisation assays to
determine absence, presence or excess expression of HJNCT and to monitor
regulation of HJNCT levels during disease therapy.

Sequence 1501 BP: 300 A: 443 C: 520 G: 238 T: 0 other:

Query Match	2.18;	Score	51.4;	DB	21;	Length	1501;
Best Local Similarity	56.1%;	Pred.	No. 0.0014;				
Matches	97;	Conservative	0;	Mismatches	76;	Indels	0;
Gaps							
QY	2296	cactgaaagccctgttgc	tgcgcacatcgagcagatgaaggagcgcagggagaagcgg	2355			
Db	1023	cccgcgagcccttgaggct	gccagcgtgagagagactcacgacaagggtgcaggacctg	1082			
QY	2356	gagctgcgcgaggtgc	ggggcgccctctctgccagtgagctggcagcggcgcctggagagt	2415			
Db	1083	gagaccgcgaatcg	gagttggagcaccagctgcggcgatgggcgcagcctggaggag	1142			
QY	2416	ggggagcctcagcaga	ggcgggccccacacagagagccacagggccaagagaggt	2458			
Db	1143	gcacggcagagcggg	agcggcgcgggctgagtgggccggcgagcgcagct	1195			
RESULT	3						
V55831							
ID	V55831	standard:	DNA; 799 BP.				
XX							
AC	V55831;						
XX							
DT	18-NOV-1998	(first entry)					
XX							
DE	Nucleotide sequence of the stabilising sequence-encoding insert.						
XX							
KW	Fusion protein; stabilising polypeptide; proteolytic degradation;						
KW	resistance; half-life; autoimmune disease; inflammation; nitro drug;						
KW	IkappaB regulator protein; inflammatory bowel disease; in vivo imaging;						
KW	nitroreductase protein; enzyme therapy; prodrug therapy; protease;						
XX	cancer; pathological condition; ss.						
XX							

V21683
ID V21683 standard; DNA; 9600 BP.
XX V21683;
XX AC
XX AC
XX AC
DT 17-AUG-1998 (first entry)
XX
XX
XX Vector plasmid pCMVMKTR-EPI.
DE
DE Polynucleotide delivery; plasmid pCMVMKTR-EPI; vector;
KW gene therapy; vaccine; polycationic agent; ss.
KW
KW Chimeric - Epstein-Barr virus.
OS Chimeric - Adeno-associated virus.
OS Chimeric - Cytomegalovirus.
OS Chimeric - Bos taurus.
XX
XX
XX Key Location/Qualifiers
DE CDS 14...2594
FT /*tag= a
FT /product= "EBV nuclear antigen A"
FT 2623..4559
FT misc_feature
FT /*tag= b
FT repeat_unit
FT 4928..5104
FT /note= "EBV origin of replication"
FT /*tag= c
FT /rpt_type= INVERTED
FT /note= "AAV inverted terminal repeat"
FT repeat_unit
FT 7189..7355
FT /*tag= d
FT /rpt_type= INVERTED
FT /note= "AAV inverted terminal repeat"
FT promoter
FT 5112..6734
FT /*tag= e
FT /note= "CMV immediate-early enhancer/promoter"
FT terminator
FT 6818..7050
FT /*tag= f
FT /note= "bovine growth hormone polyA sequence"
XX
XX W09806437-A2.
XX
XX 13-AUG-1997; 97WO-US14465.
XX
XX 13-AUG-1996; 96US-0023867.
XX
XX (CHIR) CHIRON CORP.
XX
XX Cohen F, Dubois-Stringfellow N, Dwarki V, Innis MA;
PI Murphy JE, Tetsuo U, Zukermann R;
XX
XX WPI; 1998-159296/14.
XX
XX Polycationic agents based on alpha-amino acids, able to complex
PT with nucleic acid - to facilitate its entry into cell, condense it
PT and protect it against serum degradation, particularly for use in
PT gene therapy
XX
XX Disclosure; Page 77-80; 100pp; English.
XX
XX This polynucleotide comprises the DNA sequence of vector plasmid
CC pCMVMKTR-EPI, which contains an Epstein-Barr virus (EBV) origin
CC of replication from plasmid pCEP4, a coding region for EBV nuclear
CC antigen A from pCEP4, a pair of inverted terminal repeats from
CC adeno-associated virus, a cytomegalovirus enhancer/promoter, a
CC bovine growth hormone polyA sequence, and a kanamycin resistance
CC selectable marker. Polynucleotides encoding polypeptides, such as
CC erythropoietin or leptin, and ribozymes and antisense
CC polynucleotides can be inserted into the vector. The vector is
CC preferred for use in novel compositions and methods for improved
CC polynucleotide delivery into cells. In these methods, polycationic
CC agents are used to increase the frequency of uptake of a

V21683
ID V21683 standard; DNA; 9600 BP.
XX V21683;
XX AC
XX AC
XX AC
DT 17-AUG-1998 (first entry)
XX
XX
XX Vector plasmid pCMVMKTR-EPI.
DE
DE Polynucleotide delivery; plasmid pCMVMKTR-EPI; vector;
KW gene therapy; vaccine; polycationic agent; ss.
KW
KW Chimeric - Epstein-Barr virus.
OS Chimeric - Adeno-associated virus.
OS Chimeric - Cytomegalovirus.
OS Chimeric - Bos taurus.
XX
XX
XX Key Location/Qualifiers
DE CDS 14...2594
FT /*tag= a
FT /product= "EBV nuclear antigen A"
FT 2623..4559
FT misc_feature
FT /*tag= b
FT repeat_unit
FT 4928..5104
FT /note= "EBV origin of replication"
FT /*tag= c
FT /rpt_type= INVERTED
FT /note= "AAV inverted terminal repeat"
FT repeat_unit
FT 7189..7355
FT /*tag= d
FT /rpt_type= INVERTED
FT /note= "AAV inverted terminal repeat"
FT promoter
FT 5112..6734
FT /*tag= e
FT /note= "CMV immediate-early enhancer/promoter"
FT terminator
FT 6818..7050
FT /*tag= f
FT /note= "bovine growth hormone polyA sequence"
XX
XX W09806437-A2.
XX
XX 13-AUG-1997; 97WO-US14465.
XX
XX 13-AUG-1996; 96US-0023867.
XX (CHIR) CHIRON CORP.
XX
XX Cohen F, Dubois-Stringfellow N, Dwarki V, Innis MA;
PI Murphy JE, Tetsuo U, Zukermann R;
XX
XX WPI; 1998-159296/14.
XX
XX Polycationic agents based on alpha-amino acids, able to complex
PT with nucleic acid - to facilitate its entry into cell, condense it
PT and protect it against serum degradation, particularly for use in
PT gene therapy
XX
XX Disclosure; Page 77-80; 100pp; English.
XX
XX This polynucleotide comprises the DNA sequence of vector plasmid
CC pCMVMKTR-EPI, which contains an Epstein-Barr virus (EBV) origin
CC of replication from plasmid pCEP4, a coding region for EBV nuclear
CC antigen A from pCEP4, a pair of inverted terminal repeats from
CC adeno-associated virus, a cytomegalovirus enhancer/promoter, a
CC bovine growth hormone polyA sequence, and a kanamycin resistance
CC selectable marker. Polynucleotides encoding polypeptides, such as
CC erythropoietin or leptin, and ribozymes and antisense
CC polynucleotides can be inserted into the vector. The vector is
CC preferred for use in novel compositions and methods for improved
CC polynucleotide delivery into cells. In these methods, polycationic
CC agents are used to increase the frequency of uptake of a

```

XX SQ Sequence 10380 BP; 2560 A; 2561 C; 2964 G; 2295 T; 0 other;

Query Match
Best Local Similarity 58.9%; Score 48.2; DB 20; Length 10380;
Matches 83; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 2327 agagatggagcgagcaggaggaagcggagctgcgcaggtgcggcgccctctgt 2386
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 6359 aggagcaggagggagcaggagggcgagcagcagcagcagcagcagcagcag 6418

QY 2387 ccaggagctgcagcgccctggagatggggagcctcagcagaagcgcccacacag 2446
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 6419 ggcaggagcaggagggcgagcagcagcagcagcagcagcagcagcagcagcag 6478

QY 2447 aggagccacagcccaagagg 2467
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 6479 aggggcaggagggcgagggagg 6499

RESULT 9
ID Q51731 standard; DNA; 10596 BP.
XX AC Q51731;
XX 31-MAY-1994 (first entry)
XX Plasmid pCisEBON for subcloning huHGF variants.
XX Hepatocyte Growth Factor; HGF; variant; mutein; in vitro mutagenesis;
KW proteolysis resistant; liver; malignancy; CMV-driven;
KW Cytomegalovirus; episomal expression plasmid; ss.
XX Synthetic.
XX Key
FH enhancer
FT FT Location/Qualifiers
FT FT /*tag= a
FT FT /note= "CMV enhancer/promoter"
FT FT 758..775
FT FT /*tag= b
FT FT /label= SP6_promoter
FT FT 845..849
FT FT /*tag= c
FT FT /note= "SP6 RNA start"
FT FT 902..966
FT FT /*tag= d
FT FT /function= cloning_linker
FT FT 967..1107
FT FT /*tag= e
FT FT /note= "SV40 poly A"
FT FT 1108..1531
FT FT /*tag= f
FT FT /function= SV40_origin
FT FT 1580..4189
FT FT /*tag= g
FT FT /label= EBNA-1
FT FT 4190..6374
FT FT /*tag= h
FT FT /function= orip
FT FT 4295..4887
FT FT /*tag= i
FT FT /note= "family of repeats"
FT FT 5866..5978
FT FT /*tag= j
FT FT /note= "dyad region"
FT FT 6375..6457
FT FT /*tag= k
FT FT /label= HSV_TK_terminator_3'-end
FT FT 6975..7975
FT FT /*tag= l

```

```

FT FT /phenotype= neomycin_resistance
FT FT /note= "tn5 neomycin phosphotransferase gene"
FT FT 7975..8112
FT FT /*tag= m
FT FT /label= TK_promoter
FT FT 8114..8594
FT FT /*tag= n
FT FT /function= M13_ori
FT FT 8595..10414
FT FT /*tag= o
FT FT /label= delta_2a
XX PN W092323541-A.
XX XX 25-NOV-1993.
XX PF 17-MAY-1993; 93WO-US04648.
XX XX 18-MAY-1992; 92US-0884811.
XX PR 18-MAY-1992; 92US-0885971.
XX XX (GETH ) GENENTECH INC.
XX XX Godowski PJ, Lokker NA, Mark MR;
XX XX WPI; 1993-386573/48.
XX XX Hepatocyte growth factor variants - are resistant to proteolytic
PT cleavage into its two-chain form, used to treat malignancies
PT associated with HGF receptor
XX XX Example 1; Fig 6; 87pp; English.
XX CC Plasmid pCisEBON (a pRK5 derivative) is an episomal CMV driven
CC expression plasmid. HuHGF variants with enhanced receptor binding
CC activity were produced by site-directed mutagenesis. Stable
CC populations of preferred HGF variants were obtained by transfecting
CC human embryonic kidney 293 cells and then these were subcloned in
CC pCisEBON. See R52940-R52949 for examples of pref. HGF variants.
XX XX Sequence 10596 BP; 2625 A; 2571 C; 3024 G; 2376 T; 0 other;
SQ

Query Match
Best Local Similarity 58.9%; Score 48.2; DB 14; Length 10596;
Matches 83; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 2327 agagatggagcgagcaggaggaagcggagctgcgcaggtgcggcgccctctgt 2386
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2407 aggagcaggagggagggcgagggcgagcagcagcagcagcagcagcagcag 2466

QY 2387 ccaggagctgcagcgccctggagatggggagcctcagcagaagcgcccacacag 2446
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2467 ggcaggagcaggagggagggcgagcagcagcagcagcagcagcagcagcag 2526

QY 2447 aggagccacagcccaagagg 2467
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2527 aggggcaggagggcgagggagg 2547

RESULT 10
T40348
ID T40348 standard; DNA; 10596 BP.
XX AC T40348;
XX 09-DEC-1996 (first entry)
XX Plasmid pCisEBON for expression of hepatocyte growth factor.
KW Human; hepatocyte growth factor; HGF; huHGF; serum; proteolytic cleavage;
KW pro-hormone; beta subunit; alpha subunit; kringle domain; prothrombin;
KW plasminogen; catalytic domain; serine protease; HGF variant;

```


Query Match 1.8%; Score 45; DB 19; Length 795;
Best Local Similarity 57.4%; Pred. No. 0.046;
Matches 81; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

Query Match 1.8%; Score 43.8; DB 20; Length 1528;
Best Local Similarity 53.1%; Pred. No. 0.12;

```

Query Match          1.8%; Score 43.6; DB 16; Length 217;
Best Local Similarity 55.2%; Pred. No. 0.059;
Matches 85; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

QY   69 qcaggcaaccgccgcgagcgccgcgaagaccccgtcggcaacctgcgcacgcg 128
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB   183 GCGAGCCGTCGACGGCATCAGACACC CGGTCAAGGCCGGAAGTACTCTCTCCAGAC 124
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY   129 agagaagcgcgagccgtcggggtgcttcggcgcccccaaacaccgtgtacctgcaggtggt 188
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB   123 GC GCCCGCGGATAGGGGGCGTGTTAGTCCGCCGACCCCTCAGCACGGGTCCCTCCAGGTGCT 64
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY   189 ggcagcggttagcgcggactcggcgccgcgcgcgcgc 222
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB   63 AGAAGCAGCGTCGGTGATCGGCGCGCGATCTC 30

RESULT 15
X03145
ID X03145 standard; DNA; 1567 BP.
XX AC X03145;
XX DT 23-APR-1999 (first entry)
XX DE Aspergillus terreus CBS 116.46 DNA encoding a phytase.
XX KW Phytase; modified phytase; feed additive; phytate; animal manure; ss.
XX OS Aspergillus terreus.
XX Key Location/Qualifiers
FH FT CDS 78..1530
FT FT /*tag= a
FT FT /product= phytase
FT FT /note= "contains 1 intron"
FT exon 78..124
FT FT /*tag= b
FT FT /number= 1
FT FT 125..176
FT FT /*tag= c
FT FT /number= 1
FT FT 176..1527
FT FT /*tag= d
FT FT /number= 2
XX PN A09859543-A.
XX XX
XX PD 01-OCT-1998.
XX PF 25-MAR-1998; 98AU-0059543.
XX PR 25-MAR-1997; 97EP-0810175.
XX XX
XX PA (HOFF ) HOFFMANN LA ROCHE & CO AG F.
XX PI Kostrewa D, Pasamontes L, Tomschy A, Van Loon A;
XX PI Vogel K, Wyss M;
XX XX
XX WP I; 1998-595453/51.
DR DR P-PSDB; W84357.
XX XX
XX PT New modified phytase with increased activity - produced by
PT introducing specific amino acid changes based on analysis of
PT three-dimensional enzyme structures, used as feed and food additive
XX Example 1; Fig 7; 118pp; English.
XX XX
XX CC The present sequence encodes a phytase enzyme. The specification
CC describes the production of modified phytase with improved properties.
CC This comprises modelling the three-dimensional structure of the phytase
CC being modified, and optionally that of another phytase with better
CC activity, on the basis of the 3D-structure of A. niger phytase,

```


	Query Match	1.9%	Score 48.2;	DB 3;	Length 2580;
	Best Local Similarity	58.9%;	Pred. No. 0.0021;		
	Matches 83;	Conservative 0;	Mismatches 58;	Indels 0;	Gaps 0;
Qy	2327	aggagatgaggagcgagggaggaagcggaagctgcggcagggtgcgggcgcctctctgt	2386		
Db	827	AGSAGCAGAGAGGGGCGCAGGGGCGAGGACAGGAGGAGGAGGGGCGAGGAGG	886		

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 18, 2001, 01:12:05 ; Search time 1827.78 Seconds
(without alignments)
9511.854 Million cell updates/sec

Title: US-09-434-382-1
Perfect score: 2481
Sequence: 1 atgtggcgcttgcctgcct.....agaaggtcagagccacgtga 2481

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 7991742 seqs, 3503743858 residues

Total number of hits satisfying chosen parameters: 15983484

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: em_estba:*
42: em_estfun:*
43: em_esthum1:*

44: em_esthum2:*
45: em_esthum3:*
46: em_esthum4:*
47: em_esthum5:*
48: em_esthum6:*
49: em_esthum7:*
50: em_esthum8:*
51: em_esthum9:*
52: em_esthum10:*
53: em_esthum11:*
54: em_esthum12:*
55: em_esthum13:*
56: em_esthum14:*
57: em_esthum15:*
58: em_esthum16:*
59: em_esthum17:*
60: em_esthum18:*
61: em_esthum19:*
62: em_esthum20:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estov1:*
68: em_estov2:*
69: em_estpl1:*
70: em_estpl2:*
71: em_estpl3:*
72: em_estpl4:*
73: em_estpl5:*
74: em_estro1:*
75: em_estro2:*
76: em_estro3:*
77: em_estro4:*
78: em_estro5:*
79: em_estro6:*
80: em_estro7:*
81: em_estro8:*
82: em_estro9:*
83: em_estro10:*
84: em_estro11:*
85: em_estro12:*
86: em_estro13:*
87: gb_est41:*
88: gb_est42:*
89: gb_est43:*
90: gb_est44:*
91: gb_est45:*
92: gb_est46:*
93: gb_est47:*
94: gb_est48:*
95: gb_est49:*
96: gb_est50:*
97: gb_est51:*
98: gb_est52:*
99: gb_est53:*
100: gb_est54:*
101: gb_est55:*
102: gb_est56:*
103: gb_est57:*
104: gb_est67:*
105: gb_est68:*
106: gb_est69:*
107: gb_est70:*
108: gb_est71:*
109: gb_est72:*
110: gb_est73:*
111: gb_est74:*
112: em_esthum21:*
113: em_esthum22:*
114: em_esthum23:*
115: em_estom1:*
116: em_estom2:*

117: em_estpl6:*
118: em_estpl7:*
119: em_estpl8:*
120: em_estpl9:*
121: em_estpl10:*
122: em_estpl11:*
123: em_estpl12:*
124: em_estpl13:*
125: em_estpl14:*
126: em_estpl15:*
127: em_estpl16:*
128: em_estpl17:*
129: em_estpl18:*
130: em_estpl19:*
131: em_estpl20:*
132: em_estpl21:*
133: em_estpl22:*
134: em_estpl23:*
135: em_estpl24:*
136: em_estpl25:*
137: em_estpl26:*
138: em_estpl27:*
139: em_estpl28:*
140: em_estpl29:*
141: em_estpl30:*
142: em_estpl31:*
143: em_estpl32:*
144: em_estpl33:*
145: em_estpl34:*
146: em_estpl35:*
147: em_estpl36:*
148: em_estpl37:*
149: em_estpl38:*
150: em_estpl39:*
151: em_estpl40:*
152: em_estpl41:*
153: em_estpl42:*
154: em_estpl43:*
155: em_estpl44:*
156: em_estpl45:*
157: em_estpl46:*
158: em_estpl47:*
159: em_estpl48:*
160: em_estpl49:*
161: em_estpl50:*
162: em_estpl51:*
163: em_estpl52:*
164: em_estpl53:*
165: em_estpl54:*
166: em_estpl55:*
167: em_estpl56:*
168: em_estpl57:*
169: em_estpl58:*
170: em_estpl59:*
171: em_estpl60:*
172: em_estpl61:*
173: em_estpl62:*
174: em_estpl63:*
175: em_estpl64:*
176: em_estpl65:*
177: em_estpl66:*
178: em_estpl67:*
179: em_estpl68:*
180: em_estpl69:*
181: em_estpl70:*
182: em_estpl71:*
183: em_estpl72:*
184: em_estpl73:*
185: em_estpl74:*
186: em_estpl75:*
187: em_estpl76:*
188: em_estpl77:*
189: em_estpl78:*

190: gb_gss25:*
191: gb_gss26:*
192: gb_gss27:*
193: gb_gss28:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	709.6	28.6	761	135	BE795820	BE795820 601590856
2	679	27.4	823	110	BE619259	BE619259 601473130
3	676.6	27.3	984	135	BE744876	BE744876 601576324
4	676	27.2	676	106	BE260495	BE260495 601150702
5	662.6	26.7	812	136	BE867512	BE867512 601443010
6	655.6	26.4	735	137	BE902696	BE902696 601677393
7	651.4	26.3	688	135	BE794311	BE794311 601591442
8	649	26.2	949	135	BE744197	BE744197 601577168
9	641.6	25.9	992	135	BE747163	BE747163 601577254
10	625.6	25.2	745	137	BE900936	BE900936 601674206
11	622.8	25.1	664	107	BE383336	BE383336 601298249
12	610.8	24.6	692	107	BE382353	BE382353 601298656
13	610.4	24.6	790	137	BE928933	BE928933 601435738
14	592	23.9	677	135	BE742908	BE742908 601574609
15	578	23.3	940	135	BE743831	BE743831 601577742
16	576.4	23.2	612	106	BE304720	BE304720 601106236
17	553.8	22.3	713	109	BE537086	BE537086 601063325
18	543.4	21.9	938	106	BE260626	BE260626 601146116
19	516	20.8	531	4	AA243700	AA243700 zr68g08.s
20	501.2	20.2	574	105	BE250309	BE250309 600943455
21	495.2	20.0	511	95	AW771657	AW771657 hm59h06.x
22	492	19.8	493	92	AW575677	AW575677 UI-HF-BM0
23	480	19.3	512	87	AW175923	AW175923 RCI-BR007
24	478.6	19.3	691	107	BE409312	BE409312 601300940
25	470.4	19.0	536	87	AW206103	AW206103 UI-H-BI1-
26	468.4	18.9	493	7	AA455121	AA455121 zt78c04.s
27	463	18.7	474	4	AA291670	AA291670 zt37d04.s
28	460.8	18.6	588	3	AA184645	AA184645 mt56d10.r
29	460	18.5	501	5	AA311855	AA311855 EST182568
30	451	18.2	451	17	AI200296	AI200296 qf86b12.x
31	447.2	18.0	872	110	BE615669	BE615669 601279347
32	441	17.8	855	146	W27286	W27286 28h1 Human
33	436.6	17.6	617	137	BE914535	BE914535 601656333
34	435.4	17.5	446	10	AA634909	AA634909 ab27h02.r
35	430	17.3	431	5	AA310236	AA310236 EST181085
36	429	17.3	452	10	AA679618	AA679618 ag72c12.s
37	427.2	17.2	975	105	BE250412	BE250412 600943455
38	421	17.0	421	20	AI468143	AI468143 tf92g05.x
39	404	16.3	404	88	AW304130	AW304130 xs13e05.x
40	403.6	16.3	518	144	R87541	R87541 ym89b04.r1
41	402.4	16.2	416	106	BE298273	BE298273 601118143
42	397.8	16.0	627	107	BE386924	BE386924 601274815
43	390.4	15.7	698	135	BE795434	BE795434 601592991
44	389.8	15.7	457	14	AA994126	AA994126 ou38b06.s
45	386	15.6	422	146	W37591	W37591 zc10f03.r1

ALIGNMENTS

RESULT 1
BE795820 BE795820 761 bp mRNA EST 20-SEP-2000
LOCUS 601590856f1 NIH_MGC_7 Homo sapiens cDNA clone IMAGE:3945085 5',
DEFINITION mRNA sequence.
ACCESSION BE795820
VERSION BE795820.1 GI:10217018
KEYWORDS EST.
SOURCE human.

Db	421	-TTGGAAGAGTTTCAGACCTGTCTGTGGTGGCGGCACTGCAAGCATGCGTTTGGCTGTGCGCT	479		
Qy	1959	ggtcacacctctgctggaaagtgtctattccgggggacaccatgcctgcaggctct	2018		
Db	480	GGTCACACCTCTGGCTGGAAAGTGGTCTATTCCGGGGACACATGCCCTGCGAGGCTCT	539		
Qy	2019	ggtccgagatggggaaagatgccaccctcctgatacatgaagccaccctggaagatggtt	2078		
Db	540	GGTCGGATGGGAAAGATGCACCTCTCTGATACATGAAGCCACCCCTGGAAGATGTTT	599		
Qy	2079	ggaagaagaacactgaaagaagacacagcacacaacgtcccaagccatcagctggggat	2138		
Db	600	GGAAGAGAAAGCATGTGAAAGAGACACAGCACAACTGCCAAGCCATCAGCGTGGGAT	659		
Qy	2139	gcgatatgaacggagttcattatgctgaaccacttcagccagcgtatgccaaaggctcc	2198		
Db	660	CGGATGAACGGGAGTTCATTATGCTGAACCACTTCAGC-AGGCTATGCCAAGGT-CC	716		
Qy	2199	cctttcagcccccacttcagcgagaaagtgggagttgctt	2240		
Db	717	CCTCTCAGCCCACTTCAAGCGAGAAAGTGGGAGTTGCTTT	758		
RESULT 2					
BE619259					
LOCUS	BE619259	823 bp	EST		
DEFINITION	601473130f1 NIH_MGC_68	Homo sapiens	cdna clone IMAGE:3876223 5'		
ACCESSION	BE619259	mRNA sequence.			
VERSION	BE619259.1	GI:9890197			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi				
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
TITLE	NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.				
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)				
COMMENT	Unpublished (1999)				
	Contact: Robert Strausberg, Ph.D.				
	Tel: (301) 496-1550				
	Email: Robert.Strausberg@nih.gov				
	Tissue Procurement: DCTD/DTP/Gazdar				
	cDNA Library Preparation: Life Technologies, Inc.				
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)				
	DNA Sequencing by: Incyte Genomics, Inc.				
	Clone distribution: MGC clone distribution information can be				
	found through the I.M.A.G.E. Consortium/LLNL at:				
	http://image.llnl.gov				
	Plate: LLCM623 row: k column: 08				
	High quality sequence stop: 695.				
FEATURES	Location/Qualifiers				
source	1..823				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
	/clone="IMAGE:3876223"				
	/clone_lib="NIH_MGC_68"				
	/tissue_type="large cell carcinoma"				
	/lab_host="DH10B (phage-resistant)"				
	/note="Organ: Lung; Vector: pCMV-SPORT6; Site:1: NotI; Site:2: SalI; Cloned unidirectionally. Primer: Oligo dT"				
	Average insert size 1.8 kb. Library constructed by Li Technologies."				
BASE COUNT	185 a 219 c 259 g 160 t				
ORIGIN					
Query Match 27.4%; Score 679; DB 110; Length 823;					
Best Local Similarity 92.7%; Pred. No. 4.7e-167;					
Matches 757; Conservative 0; Mismatches 55; Indels 5; Gaps					
Qy	1562	ttggcagctgtcgcttattcagagaccaggtggacaggtgtcctggcaccctgctg	1621		


```

QY 1282 cgtccaggaggagtgccagagagatgccattattacttgaatcctgagggaattcata 1341
Db 609 COTCCAGAGGAGGTGGCAGAGGATGCCATTATTACTTGCAATCCTGAGGAATTCATA 668
QY 1342 gttgaggcgtcagcttcccaactccagcagagcgtgacagaggtacagagagagtgcg 1401
Db 669 GTTAGGCGCTG-AGCTTCCCAACTTCCAGCAGAGCGTGCAGAGTACAGAGGAGTGCG 727
QY 1402 caggacgccccagccccag 1420
Db 728 CAGGAGGCCAGCCAGCAG 746

RESULT 4
BE260495 676 bp mRNA EST 13-JUL-2000
LOCUS 601150702F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:3503184 5',
DEFINITION mRNA sequence.
ACCESSION BE260495
VERSION BE260495.1 GI:9131807
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 676)
AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LUCM176 row: d column: 01
High quality sequence stop: 672.
FEATURES
    Location/Qualifiers
        1..676
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:3503184"
            /clone_lib="NIH_MGC_19"
            /tissue_type="neuroblastoma"
            /lab_host="DH10B (phage-resistant)"
            /note="Organ: brain; Vector: pORF7; Site_1: XhoI; Site_2:
            EcoRI; cDNA made by oligo-dT priming. Directionally
            cloned into EcoRI/XhoI sites using the following 5',
            adaptor: GGCACGAG(G). Library constructed by Ling Hong
            in the laboratory of Gerald M. Rubin (University of
            California, Berkeley) using ZAP-cDNA synthesis kit
            (Stratagene) and Superscript II RT (Life Technologies).
            Note: This is a NIH_MGC Library."
BASE COUNT 154 a 207 c 176 g 139 t
ORIGIN

```

```

Query Match 27.2% Score 676; DB 106; Length 676;
Best Local Similarity 100.0%; Pred. No. 2.7e-166;
Matches 676; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1093 accagcacttgctgaatgagaactgctcagttcacaacttcgagccacaag 1152
Db 1 ACCGAGCACTTGGTCTGAATGAGAACTGTGCTCAGTTCAACACTTCGACGCCACAAG 60
QY 1153 attcaaacccagctcaacctctaccaccggacatcttccccctgctcaccagtttccgc 1212
Db 61 ATTCAAAACCCAGCTCAACCTCATCCACCGGACATCTTCCCTGCTCACCAGTTTCCGC 120
QY 1213 tgtaagaagaggagccccaccctcagtgtycccatggttcagggtgaatgcctctcaag 1272

```

```

Db 121 TGTAAAGAGAGGGCCCCACCCCTCAGTGTGCCCATGGTTACAGGTGAATGCCCTCAAG 180
QY 1273 taccagctccgtcccgaggagtgccagagagtgccattattacttgcaatcctgag 1332
Db 181 TACCAGCTCCGTCCCGAGGAGGTGGCAGAGGATGCCATTATTACTTGCATCTCTGAG 240
QY 1333 gaattcatagtgagggcgtgcagcttcccaacttcacagcagagcgtgcaggagtagcag 1392
Db 241 GAATTCATAGTTGAGCGCTGCAGCTTCCCAACTTCCAGCAGAGCGTGCAGGAGTACAGG 300
QY 1393 aggagtgccagagagccccagccccagcagagaaagaagtcagtaaccagaaaatc 1452
Db 301 AGGAGTGCAGAGAGCGGCCAGCCCGCAGAGAGAAAGAGTACAGTACCCAGAAATCATC 360
QY 1453 ttcccttggaacagggtctgccatcccgatgaagattcgaattcgaatgctcagtgccacattgtc 1512
Db 361 TTCTCTTGGAACAGGGTCTGCCATCCGCATGAAGATTTCGAATGTCAGTGCACACTTGTG 420
QY 1513 aacataagccccgacacgctctctgctactggactggtgaggggcacatttggggcagctg 1572
Db 421 AACATAAGCCCCGACACAGCTCTCTGCTACTGGACTGTGGTGCAGGCGACATTTGGCGAGCTG 480
QY 1573 tgccgtcatcaggaagacagagtgacaggggtctctgggacccctggcgtgctgtgtgtg 1632
Db 481 TGCCGTCTATTAGCGAGACAGGTGGACAGGGTCTCTGGCACCCCTGGCTGCTGTGTGTGTG 540
QY 1633 tccacactcagcagagatcacacacggcttgccaggtatcttctgctcagagagaacgc 1692
Db 541 TCCACCTCAGCGAGATCACCACGGGCTTGCCCAAGTATCTTGTCTGAGAGAGAACGC 600
QY 1693 gccttggcatctttgggaaagccgcttcaccccttctgctggtggtgttcccccaaccagctc 1752
Db 601 GCCTTGGCATCTTTGGGAAAGCGCTTACCCCTTCTGCTGTGTTGCCCCCAACCAAGCTC 660
QY 1753 aaagccttgctccagc 1768
Db 661 AAAGCCTGGCTCCAGC 676

RESULT 5
BE867512 812 bp mRNA EST 27-SEP-2000
LOCUS 601443010F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3847226 5',
DEFINITION mRNA sequence.
ACCESSION BE867512
VERSION BE867512.1 GI:10316288
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 812)
AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM548 row: c column: 03
High quality sequence stop: 686.
FEATURES
    Location/Qualifiers
        1..812
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:3847226"

```

/clone_lib="NIH_MGC_65"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.8 kb. Library constructed by Life
Technologies."

BASE COUNT 200 a 216 c 220 g 176 t

Query Match 26.7%; Score 662.6; DB 136; Length 812;
Best Local Similarity 97.4%; Pred. No. 9.4e-163;
Matches 716; Conservative 0; Mismatches 14; Indels 5; Gaps 4;

Qy 724 gctttcattctgaagcttcacttaagagagagaaactcttgggtgctcaaaagcaagagag 783
Db 1 GCTTTCATCTGTAAGCTTCACCTTAAGAGAGGAACCTCTTGGTGCTCAAGCAAGGAG 60
Qy 784 atg9g9cctccagttggagcagctgcccacgctcccatcattgctgtgtcaagagacggg 843
Db 61 ATGGGCTCCCGAGTTGGGACAGCTGCCATCGCTCCCATCATTTGCTGTCAAGGACGGG 120
Qy 844 aaagcatcactcatgaaggaagagagatttggctgaagagctgtgtactcctccagat 903
Db 121 AAAGCATCATCTATGAAGAGAGAGAGATTGGCTGAAGAGCTGTGTACTCTCCAGAT 180
Qy 904 cctggtgctgttttgggtgtagaattccagatgaagcttcattcaaccatctgt 963
Db 181 CTTGCTGTGC-TTTGTGGTGAAGATGCCAGATGAAGCTTCATTCACCCATCTGT 239
Qy 964 gagaatgccactttcagaggtacaaaggaagcagatgccccgctggtcgttgggtt -1023
Db 240 GAGAATGCCACTTTACAGAGGTACCAAGAAAGGACAGATGCCCGCTGGCTGTGGT 299
Qy 1024 cacatggccagcatctgtctgttgacagcaggtaccagagtgatgagagagttt 1083
Db 300 CACATGGCCCCAGCATCTGTCTGTGACAGCAGGTACCAGCAGTGGATGGAGAGGTTT 359
Qy 1084 gggcctgacacccagcacttggctgtaagtgaactgtgctcagttcacaaccttcgc 1143
Db 360 GGGCTGACACCCAGCACTTGGTCTGAATGAGAACTGTGGCTCAGTTCAACACCTTCGC 419
Qy 1144 agccacaagtatcaaacccagctcaacatccacccggacatcttccccctgtcacc 1203
Db 420 AGCCACAAGATTCAAAACCCAGCTCAACCTCATCCACCCGGACATCTTCCCTGTCTACC 479
Qy 1204 agttccgctataaagagagagcccccacccctcagtgcccatggttcagggtgaatgc 1263
Db 480 AGTTTCCGCTGTAAAGAGAGGGCCCCACCTCTCAGTGTGCCCATGGTTTCAGGGTGAATGC 539
Qy 1264 ctctcaagtaccagctcgtcccagagggg--agtggcagagggatgccattattact- 1320
Db 540 CTCCTCAAGTACCAGCTCCGTCCTCCAGGAGGCGAGTGGCGAGGGATGCCATTATTACTC 599
Qy 1321 tgcatactcagggaattcaatagtgagcgtcagcttcccaactccacagcagagcgtg 1380
Db 600 TGCAATCTGAGGAATTATCATATTGAGGCGCTGCAGCTTCCCAACTT-CAGCAGAGCGTG 658
Qy 1381 cagagttacagagagagtgccagagcggccagcccccagcagagaaaaagagcagttac 1440
Db 659 CAGGAGTACAGGAGAGTGGCCAGGAGCGGCGCCAGCCAGCAGAGAGAAAGATCAGTAC 718
Qy 1441 ccagaaatcatcttc 1455
Db 719 CCAGAAATCATCTCC 733

RESULT 6
BE902696 735 bp mRNA EST 29-SEP-2000
LOCUS 601677393F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:3959926 5',
DEFINITION mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

BE902696
BE902696.1 GI:10393148
EST.
human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 735)
NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.

National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov

Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov

Plate: L1CM841 row: j column: 23
High quality sequence stop: 732.
Location/Qualifiers
1..735

source

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3959926"
/clone_lib="NIH_MGC_21"
/tissue_type="choriocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: placenta; Vector: pOTB7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCAGGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-CDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 168 a 206 c 203 g 157 t

1 others

Query Match 26.4%; Score 655.6; DB 137; Length 735;
Best Local Similarity 97.4%; Pred. No. 6.2e-161;
Matches 719; Conservative 0; Mismatches 14; Indels 5; Gaps 5;

Qy 1190 tccccctgctcaccagtttccgctgtaagaagagggggcccccaccctcagtgcccatgg 1249
Db 1 TCCCCCTGCTCACCAGTTTCCCCGTGTAAAGAGGAGGGCCCCCCTCAGTGTGCCATGG 60
Qy 1250 ttcaggggtgaatgctcctcactcaggtaccagtcctccagagggagtgccagagggatg 1309
Db 61 TTCAGGATGAATGCTCCTCAAGTACCAGTCCGTCCTCCAGGAGGAGTGGCAGAGGGATG 120
Qy 1310 ccatattactt-gcaatcctcagggaattcatagttgagcgtcagcttcccaacttc 1368
Db 121 CCATTATTACTTNGCAATCCCTCAGGAATTCATAGTTGAGGCGCTGCAGCTTCCCAACTTC 180
Qy 1369 cagcagagcgtgagaggtacagagagagtgctcagagagcggcccccagcagagagaaaa 1428
Db 181 CAGCAGAGCGTGCAGGAGTACAGGAGGAGTGCAGGAGGCGGCCAGGCCAGCAGAGAGAAA 240
Qy 1429 agaagtcagtcaccagaaaatcatcttcttggaaacaggtctgcctcctcagatgaagatt 1488
Db 241 AGAAGTCAGTACCAGAAAATCATCTTCTTGGACAGGGTCTGCATCCCCGATGAAGATT 300
Qy 1489 cgaaatgicagtgccacacttgtcaacataagccccacacgtctctctgctactgactgt 1548
Db 301 CGAAATGTCTAGTGGCACACTTGTCAACATAAGCCCCGACACGTCTCTCTACTTGACTGT 360
Qy 1549 ggtgagggcacatttggcagctgtgccgttcattaccagagaccaggtggacagaggtcctg 1608
Db 361 GGTGAGGCGACATTTGGGACGCTGTGCGCTCATTTACGGAGACACAGGTGGAGGAGGTCCTG 420

Tissue Procurement: DCTD/DTP
 CDNA Library Preparation: Ling Hong/Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
 plate: LLC524 row: 9 column: 18
 High quality sequence stop: 668.

FEATURES

source
 1. .949
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3838121"
 /clone_lib="NIH_MGC_9"
 /tissue_type="adenocarcinoma cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: ovary; Vector: pOTB7; Site_1: XhoI; Site_2:
 EcoRI; cDNA made by oligo-dT priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5'
 adaptor: GGCACGAG(G). Size-selected >500bp for average
 insert size 1.8kb. Library constructed by Ling Hong in
 the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies)."
 BASE COUNT 225 a 263 c 281 g 179 t 1 others
 ORIGIN
 Query Match 26.2%; Score 649; DB 135; Length 949;
 Best Local Similarity 92.6%; Pred. No. 3.6e-159;
 Matches 773; Conservative 0; Mismatches 45; Indels 17; Gaps 8;
 QY 672 tccacatggtgttagccagagaaggggtcagggaacttccctggtcgtagcttcat 731
 DB 1 TCCACATGGTGTAGCCAGAGAGAGGGGTGAGGACTTCCCTGCTGTAGCTTCAT 60
 QY 732 ctgtaagcttcaactaaagagagaactctgtgtctcaagcaagagatggcct 791
 DB 61 CTGTAAGCTTCACTTAAAGAGAGAACTCTTGGTGCTCAAAAGCAAGAGATGGCCT 120
 QY 792 cccagttggagacagctgccatcgctccatcattgctgctcaagacgggaaagcat 851
 DB 121 CCCAGTTGGACAGCTGCCATCGCTCCCATCATTTGCTGTCAAGACGGGAAACCAT 180
 QY 852 cactcatgaaggaagagatgttggctgaagagctgttactctccagatcctgtgc 911
 DB 181 CACTCATGAAGGAAGAGAGA-TTTGGCTGAAGAGCTGTACTCTCCAGATCCTGGTGC 239
 QY 912 tgcctttgtgtgtaaatgtccagatgaagcttcatcaacccatctgtgagaatgc 971
 DB 240 TGC-TTTGTGGTGTAGAAATGTCAGATGAAGCTTCATTCAACCCCATCTGTGAGATGC 298
 QY 972 cacccttcagaggtaccagaagaaggcagatgccccgtggccttgggttccatcagtc 1031
 DB 299 CACCTTTCAGAGGTACCAAGGAAGGAGATGCCCGCTGGCTTGGTGTTCATATGC 358
 QY 1032 cccagcatctgtctgtgtgacagcaggtaccagcagctgagagaggtttgggcctga 1091
 DB 359 CCCAGCATCTGTGCTTGTGACACAGCAGGTACCAGCAGTGGATGGAGAGGTTTGGCCCTGA 418
 QY 1092 caccagcaacttggtcctgaatgagaactgtgctcagttcacaaccccttcgcagccaca 1151
 DB 419 CACCCAGCACTTGGCTGAATGAGAACTGTGCTCAGTTCAACAACCTTCGCAGCCACAA 478
 QY 1152 gattcaaacccagctcaacctcatccaccggagcatcttcccctgctcaccagtttccg 1211
 DB 479 GATTCAAACCCAGCTCAACCTCATCCACCGGACATCTCCCTGCTCACCAGTTTCCG 538
 QY 1212 ctgtaagaagagggggcccccacctcagtggtgcccattggttccaggtgaaatgcctctcaa 1271
 DB 539 CTGTAAGAAGAGAGGGGGCCCCACCTCAGTGTGCCATGGTTCAGGGGTGAATGCCCTCTCAA 598
 QY 1272 gtaccagctccgtcccccagaggaggtggcagagggatgccattattac-ttgcaactctg 1330

Db 599 GTACAGCTCGT-CCAGGAGGGAGTGGCAGAGGATGCCATTATTCTTTGCAATCCTG 657
 QY 1331 aggaattcatagttgagcgtcgcagc-----ttcccaactccagca-gagcgtgcag 1383
 Db 658 AGGAATTCATAGCTTGAGGCGGGTGCAGGCTTCCCAAACTCCAGCAGAGGCTGCAG 717
 QY 1384 ---gagtcacagggaggtgctgcaggacgcccagcccccag---cagagaaagaagtcag 1437
 Db 718 GGAGTACCCGAGGAGTGCAGCAGGCGCCAGCCAGCCAGCAGCAAGAAAGTCAG 777
 QY 1438 taccagaatcatcttcccttggaacaggggtctgcatcccgatgaagattcgaa 1492
 Db 778 GCACCAAAATCACTTCTCTGGAACAGGCTTGCACCCCGGATGAAGATCGAAA 832

RESULT 9

BE747163 992 bp mRNA EST 15-SEP-2000
 LOCUS 601577254F1 NIH_MGC_9 Homo sapiens cDNA clone IMAGE:3838237 5',
 DEFINITION mRNA sequence.
 ACCESSION BE747163
 VERSION BE747163.1 GI:10161155
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 CONTACT: Robert Strausberg, Ph.D.
 TEL: (301) 496-1550
 EMAIL: Robert.Strausberg@nih.gov
 TISSUE Procurement: DCTD/DTP
 CDNA Library Preparation: Ling Hong/Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
 plate: LLC524 row: 1 column: 14
 High quality sequence stop: 781.

FEATURES

source
 1. .992
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3838237"
 /clone_lib="NIH_MGC_9"
 /tissue_type="adenocarcinoma cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: ovary; Vector: pOTB7; Site_1: XhoI; Site_2:
 EcoRI; cDNA made by oligo-dT priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5'
 adaptor: GGCACGAG(G). Size-selected >500bp for average
 insert size 1.8kb. Library constructed by Ling Hong in
 the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies)."
 BASE COUNT 225 a 287 c 313 g 167 t
 ORIGIN

Query Match 25.9%; Score 641.6; DB 135; Length 992;
 Best Local Similarity 97.9%; Pred. No. 3.2e-157;
 Matches 693; Conservative 0; Mismatches 9; Indels 6; Gaps 4;

QY 991 ggaagagcagatccccctggccttggtgttcaatgcccccagcatctgtctgtg 1050
 Db 1 GGAAGGACAGATGCCCGTGGCC-TGGTGGTTACATGGCCCGCAGCATCTGTCTGTG 59
 QY 1051 gacagcaggtaccagcagtgatggagaggttggcctgacacccagcattgctgtc 1110


```

RESULT 11
BE383336      664 bp      mRNA      EST      21-JUL-2000
LOCUS      601298249F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:3628308 5',
DEFINITION      mRNA sequence.
ACCESSION      BE383336
VERSION      BE383336
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 664)
NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LLC313 row: a column: 13
High quality sequence stop: 662.
FEATURES
Location/Qualifiers
1..664
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3628308"
/clone_lib="NIH_MGC_19"
/tissue_type="neuroblastoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5',
adaptor: GGACAGAG(G). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT      154 a 199 c 174 g 137 t
ORIGIN
Query Match      25.1%; Score 622.8; DB 107; Length 664;
Best Local Similarity 99.4%; Pred. No. 2.4e-152;
Matches 646; Conservative 0; Mismatches 2; Indels 2; Gaps 2;
QY 1093 accagcactgttcctgaatgagaactgtgcctcaggtcacacacctgcagccacaag 1152
|
|
|
Db 1 ACCAGCAGCTTGTCTGATGAGAACTGTGCCTCAGTTCACACACTTCCAGCCACAAG 60
|
|
|
QY 1153 attcaaacccagctcaacctcattccaccggagacttccccctgtcaccagtttcgc 1212
|
|
|
Db 61 ATTCAAAACCCAGCTCAACCTCATCCACCCGAGATCTTCCCTCCGTCACAGTTCCGC 120
|
|
|
QY 1213 tgaagaagagagggccccaccctcagtggtccatgggtcagggtgaatccctcctcaag 1272
|
|
|
Db 121 TGTAAAGAGAGGGGCCCCACCCCTCAGTGTGCCCATGTTCCAGGTGAATGCCCTCTCAAG 180
|
|
|
QY 1273 taccagctccgtccacagagggagtgagcagagggatgccattattacttcaatcctgag 1332
|
|
|
Db 181 TACCAGCTCCGTCACAGGAGGAGTGGCAGAGGATGCCATTTATTACTTCCATCTCGAG 240
|
|
|
QY 1333 gaattcatagttgagcgctgcagcttcccaacttcacagcagcgtgcaggagtagcag 1392
|
|
|
Db 241 GAATTTCATAGTTGAGGCGCTGCAGCTTCCCAACTTCCACAGAGCGTGCAGGATACAGG 300
|
|
|
QY 1393 aggagtgcgcaggacgccccagccccagagagaaaaagaagtacgtaccaccagaaatcatc 1452
|
|
|

```

```

Db 301 AGGAGTCGCGAGGACGCCAGCCAGCCAGAGAGAGAGTACAGTACCCAGAGAAATCATC 360
|
|
|
QY 1453 ttcttgaacagaggtctgccatcccccgaatgaattcgaattcgaattcgaattcgaattcgc 1512
|
|
|
Db 361 TTCCTTGAACAGAGGTCTGCCATCCCGATGAAGATTGCAATGTCAGTGCACACTTGTG 420
|
|
|
QY 1513 aaataagccccacacagctctctgctactgactgtgtgagggcacatttgggcagctg 1572
|
|
|
Db 421 AACATAAGCCCCGACACAGTCTCTGCTACTGTGAGTGTGTGAGGSCACATTGGGCAGCTG 480
|
|
|
QY 1573 tgcgctcattacagagaccaggttgacagaggtctctggccacccttgctgtgtgtttgtg 1632
|
|
|
Db 481 TGCCTGCTATTACGAGAGACCAAGGTGGACAGGGTCTCTGGGCACCCCTGGTGTGTGTTGTG 540
|
|
|
QY 1633 tcccacctgcacgacagatcacacacgggcttgcacagttatctgtgc-agagagaacg 1691
|
|
|
Db 541 TCCACCTGCACGAGATCACACACGGGCTTGCCAGTAGTATCTTGTCTGCAAGAGAGAACG 600
|
|
|
QY 1692 cgcctggcatctttgggaagccgcttcaccccttgcctgtgtgtgtgtgtgtgtgtgt 1741
|
|
|
Db 601 CGCCTTGGCATCTTTGGGAAAG-CGCTTCACCCCTTCTGCTGTGTGTGTGTGTGTGTGT 649
|
|
|
RESULT 12
BE382353      692 bp      mRNA      EST      21-JUL-2000
LOCUS      601298656F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:3629028 5',
DEFINITION      mRNA sequence.
ACCESSION      BE382353
VERSION      BE382353
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 692)
NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LLC314 row: o column: 13
High quality sequence stop: 600.
FEATURES
Location/Qualifiers
1..692
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3629028"
/clone_lib="NIH_MGC_19"
/tissue_type="neuroblastoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5',
adaptor: GGACAGAG(G). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT      152 a 193 c 197 g 150 t
ORIGIN
Query Match      24.6%; Score 610.8; DB 107; Length 692;
Best Local Similarity 98.8%; Pred. No. 3.4e-149;

```

Matches	637;	Conservative	0;	Mismatches	7;	Indels	2;	Gaps	2
QY	1475	tccgatgaagatcgaataatcagtcgacacacttgccaacataagccccgacacgctctc	1534						
Db	1	TCCCGATGAAGATTTCGAAATGTCAGTGCACACATTGTCAACATTAAGCCCCGACAGTCTC	60						
QY	1535	tgcctactggagctgtggtgagggacacatttggcgagctgtgcccgtattacagagaccagg	1594						
Db	61	TGCTACTTGGACTGTGGTGAAGGCAATTTGGCGACGCTGCGCGTCATTACGGAGACCAGG	120						
QY	1595	tggacaaggctcctgggacccctggctgctgtgtttgtgtccaccctgcacagatcacc	1654						
Db	121	TGGACAGGCTCTGGGACCCCTGGCTGTGTGTGTGTGTCACCTGCACGACAGATACC	180						
QY	1655	acacgggcttgccaaagtattctgtgcagagagaacgccttggcattctttgggaaagc	1714						
Db	181	ACACGGGCTTGCCAAATATCTTGTGCAGAGAGAACGCGCTTGGCATCTTTGGGAAGC	240						
QY	1715	cgcttcacccttctgctgggtgtgttgcctcccaaccagctcaaaagctgcctcagctacc	1774						
Db	241	CGCTTCAACCTTTGCTGGTGTGTGTCGCCCAACACAGCTCAAAGCCTGGCTCCAGCAGTACC	300						
QY	1775	acaaccagtcgagagaggtcctgcacacatcagtatgattccctgccaatgccttcagg	1834						
Db	301	ACAAACAGTCGACGAGGTCTCTGCACACACATCAGTATGATTTCTTGCCAAATGCCCTTCAGG	360						
QY	1835	aagggctggaatcctccagctcctcagtggaagatgtagcttcgctgttgcaaacat	1894						
Db	361	AAGGGCTGAGATCTCCAGTCTCGACTGGGAAGATTGATCAGTTCGCTGTTCGAAACAT	420						
QY	1895	gtgatttggaaagatttcagacctgtgtgtgcgcactgcaagcatcgcttggctgtg	1954						
Db	421	GTGATTTGGAAGAGTTTCAGACCTGTCTGTGTGCGGACTGCAAGCATGCGTTTGGCTGTG	480						
QY	1955	cgctgggtgcacacctctgctggaaagtgtgtctattccggggacacacatgccttcgcagg	2014						
Db	481	CGCTGGTGCACACCTCTGGCTGGAAAGTGGTCTATTCGGGGGACACCATGCCCTGCGAGG	540						
QY	2015	ctctggttcggatggggaagatgccacctctctgatatcatgaagccaccctggaagatg	2074						
Db	541	CTCTGGTCCGATGGGNAAGATGCCACCTCTCTGTATCATGTAAAGCCA-CTTGGAAAGATG	599						
QY	2075	gtttggaaggaagcagctgtggaagacacacacagcacacgctccca	2120						
Db	600	GTTTGGCAGAGGAAGCCGTGG-AAAGACACACAGCACACGTCCTCCAA	644						
RESULT	13								
BE892893									
LOCUS	601435738F1 NIH_MGC_72	790 bp	mRNA	EST	29-SEP-2000				
DEFINITION	Homo sapiens cDNA clone IMAGE:3920792 5', mRNA sequence.								
ACCESSION	BE892893								
VERSION	BE892893.1	GI:10353525							
KEYWORDS	EST.								
SOURCE	human.								
ORGANISM	Homo sapiens								
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;								
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.								
TITLE	1 (bases 1 to 790)								
JOURNAL	NH-MGC http://www.ncbi.nlm.nih.gov/MGC/ .								
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)								
	Contact: Robert Strausberg, Ph.D.								
	Tel: (301) 496-1550								
	Email: Robert_Strausberg@nih.gov								
	Tissue Procurement: ATCC/DCPD/DTF								
	CDNA Library Preparation: Life Technologies, Inc.								
	CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)								
	CDNA Sequencing by: Incyte Genomics, Inc.								
	Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov								

Plate: LLCW739		row: 1	column: 09		
High quality sequence stop: 662.					
FEATURES	Location/Qualifiers				
	1..790				
source	/organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:3920792" /clone_lib="NTH_MGC_72" /tissue_type="melanotic mel" /lab_host="DH10B (phage-res /note="Organ: skin; Vector: Site2: SalI; Cloned unidir Average insert size 2 kb. Technologies."				
BASE COUNT	185 a	221 c	220 g 164 t		
ORIGIN					
Query Match					
Best Local Similarity		24.6%;	Score 610.4;		
Matches 726;		Conservative	93.6%; Pred. No. 4.5e		
		0; Mismatches			
QY	903	tctcgtgtcgtcttttgggtggtgagatgtccagatg			
Db	1	TCCTGTGTGCTTGTGGTGGTGAAGATGTCACAGATG			
QY	963	tgagaatgccacctttcagaggtaccaggaagaagcag			
Db	61	TGAGATGCCACCTTTCAGAGGTACCAAGGAAGGGCAG			
QY	1023	tcacatggccccagcatctgtctgtgtggacagcaggt			
Db	121	TCACATGGCCCCAGCATCTGTGCTTGTGGACAGCAGGT			
QY	1083	tggcctcgcacacccagcacttggctcctgaaatgagaact			
Db	181	TGGGCGCTGACACCCAGCAGCTTGGTCTCTGAATGAGAACT			
QY	1143	cagccacaagattcaaacceagctcaacctcatccacc			
Db	241	CAGCCACAAGATTCAAAACCCAGCTCAACCTCAATCCACCC			
QY	1203	cagtttcgcgtgtaagaaggaggggccccaccctcagtg			
Db	301	CAGTTTCCGCTGTAAAGAGAGGGGCCCCACCCTCAGTGT			
QY	1262	gcctcctcaagtaccagctccgtccccagagaggagtgg			
Db	361	GCCTCTCAAGTACCAGCTCCGTCCACGAGGGAGTGGG			
QY	1322	gcaatcctgaggaattcatagttgaggcgtgcagcttt			
Db	421	GCAATCCTGAGGAATTATAGTTTGAGGGCGCTGCAGCTTT			
QY	1381	caggaggtacaggagaggtgcgcaggacgcgccagcccc			
Db	481	CAGGAGTACAGGAGGAGTGCAGAGACGCGCCACAGCCCA			
QY	1440	cccagaaatcatctctcttggaaacagggtctgcocatccc			
Db	541	CCCAGAAATCATCTTCCTTGGAAACAGGGTCTGCCATCCC			
QY	1498	agtgccacacattg---tcaacataagcccccgacacagtc			
Db	601	AGTGCCACACATTGGTCTCATATAAAGCCCCCGAACACGCTC			
QY	1555	ggcacattt-gggcagctgtgccgttcattacgg-agac			
Db	661	GGCACATTTGGGGCAAGTGTGGCGGCATTACGGGAAGAC			
QY	1613	ccctggcgtgct---gtgtttgtgtcccaacctgcacgc			
Db	721	CACCTGGGCTGGCGGGTTTGGTTTCCACCGCGCAGCGC			

```

RESULT 14
LOCUS BE742908 677 bp mRNA EST 15-SEP-2000
DEFINITION 601574609F1 NIH_MGC_9 Homo sapiens cDNA clone IMAGE:3835658 5',
mRNA sequence.
ACCESSION BE742908
VERSION BE742908.1 GI:10156900
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 677)
AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: DCTD/DTP
cDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LLCM518 row: a column: 03
High quality sequence stop: 672.
Location/Qualifiers
1..677
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3835658"
/clone_lib="NIH_MGC_9"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
BASE COUNT 160 a 173 c 217 g 127 t
ORIGIN
Query Match 23.9%; Score 592; DB 135; Length 677;
Best Local Similarity 99.0%; Pred. No. 2.9e-144;
Matches 669; Conservative 0; Mismatches 0; Indels 7; Gaps 7;
QY 1807 agtatgattcctcgaatgccttcagaaaggctagatctcagtcctgcagtgagaa 1866
Db 1 AGTATGATTCCTGCCAATGCTTCAGGAAGGGCTGAGATCTCCAGTCTGCAGTGGAA 60
QY 1867 agattgatcattcctgttgcgaacatgtgatttggaaagatttcagacctgtcgtg 1926
Db 61 AGATTGATCAGTTCGCTGCTTCGGAACATGTGATTGGGAAGATTTCAGACCTGCTG 120
QY 1927 cggcaactgcaagcatcgcttggctgtgcgtgtgcacacctctggtggaagtggtc 1986
Db 121 CGGCACCTGCAAGCATGCGCTTGGCTGTGCGCTGGTGTCACACCTCTGGCTGGAAGTGGTC 180
QY 1987 tattccgggaacacatgccttcgagagctctgctcggatgggaaagatgcacacctc 2046
Db 181 TATTCCGGGACACCATGCTCCCTGGAGGCTCTGCTCGGATGGGGAAGATGCCACCTC 240
QY 2047 ctgatacatgaagccaccttggaagatggtttggaaggaagcagtggaagaaacacac 2106
Db 241 CTGATACATGAAGCCACCTTGAAGATGG-TTGAAGAGGAAGCAGTGGAAAGACACAC 299

```

```

QY 2107 agcaaacgttcccaagccatcagcgtggggatgggatgaacgcggagttcattatgctg 2166
Db 300 AGCACACGTCCCAAGCCATCAGCGTGGGATGCGGATGAACGCGAGTTCATTATGCTG 359
QY 2167 aaccattcagccagcgtatgccaagtccttcagcccccaacttcagcgagaaa 2226
Db 360 AACCACCTTCAGCCAGCGCTATGCCAAGGTCCCTCTTCAGCCCCCACTTCAGCGAGAAA 419
QY 2227 gtggagttgcttttgaccacatgaagtcgtcttggagactttcccaaatgcccaag 2286
Db 420 GTGGAGTTGCTTTGACCACATGAAGCTGCTG-TTGGAGACTTTCCAAACATGCCCAAG 478
QY 2287 ctgattcccccaactgaaagccc-tgttctggtggacatcagaggatgaggagcgcag 2345
Db 479 CTGATTCCTCCACTGAAAGCCCGTGTTCCTGGCGACATCGAGGAGATGGAGAGCGCAG 538
QY 2346 ggagaagcggagcgtcggaaggtcgggcgccctctctgtccagggagcgtggaaggcg 2405
Db 539 GGAGAAGCGGAGCTGCGGAGGTGCGGGCGG-CCTCCTGTCCA-GGAGCTGGCAGGCGG 596
QY 2406 cctggaggtgggagcctcagcagaagcggccccacacagaggagccacagggcccaagaa 2465
Db 597 -CTGGAGATGGGAGCCTCAGCAGAGAGCGGG-CCACACAGAGGAGCCACAGGCCAAGAA 654
QY 2466 ggtcagagcccaagtga 2481
Db 655 GGTGAGAGCCCAAGTGA 670
RESULT 15
LOCUS BE743831 940 bp mRNA EST 15-SEP-2000
DEFINITION 60157742F1 NIH_MGC_9 Homo sapiens cDNA clone IMAGE:3926362 5',
mRNA sequence.
ACCESSION BE743831
VERSION BE743831.1 GI:10157823
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 940)
AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: DCTD/DTP
cDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LLCM54 row: d column: 11
High quality sequence stop: 631.
Location/Qualifiers
1..940
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3926362"
/clone_lib="NIH_MGC_9"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
BASE COUNT 250 a 242 c 281 g 167 t

```


ORIGIN

Query Match. 23.3%; Score 578; DB 135; Length 940;
Best Local Similarity 89.5%; Pred. No. 1.5e-140;
Matches 725; Conservative 0; Mismatches 70; Indels 15; Gaps 9;

QY 672 tccacatggtgttagccagagaaaggggtcagggaactctccctggtgtagcttcat 731
Db 1 TCCACATGGTGTAGCCAGAGAGAGGGGTCA-GGACTCTTCCCTGGTGTAGCTTTTCAT 59

QY 732 ctgtaagcttcaactaaagagagaaactcttctgctcaagcaaaagagatggcct 791
Db 60 CTGTAAGCTTCACTTAAGAGAGAGAAACTTCTTGGTGTCAAGCAAAAGAGATGGCCT 119

QY 792 cccagttgggacagctgcccacatcattctgctgtaaggacgggaaaaagcat 851
Db 120 CCCAGTTGGGACAGCTGCCATCGCTCCCATCATTTGCTGTGTCAGGACGGGAAAAAGCAT 179

QY 852 cactcatgaaggaagagagattttggctgaagagctgtgtaactctccagatccctggctgc 911
Db 180 CACTCATGAAGGAAGAGAGATTTTGGCTGAAGAGCTGTACTCTCCAGATCTCTGGTGC 239

QY 912 tgcttttgggtgtagaatgccagatgaagcttcatcaccatctgtgaaatgc 971
Db 240 TGCTTTTGGTGTAGAAATGTCCAGATGAAGACTTTCATTAACCCATCTGTGAGAAATGC 299

QY 972 caccittcagaggtaccaggaagagcagatgcccccggtggccttgggttcacatgac 1031
Db 300 CACCTTTTCAGAGGTACCAAGGAAGAGCAGATCCCCCGTGGCTTGGTGTTCACATGG- 358

QY 1032 cccagcatctgtctgttgagcagcaggtaccagcagtgatggagaggtttggcctga 1091
Db 359 CCCAGCATCTGTGCTTGTGGAGCAGCAGGTACCAGCAGTGGATGGAGAGG-TTGGGCTGA 417

QY 1092 caccagcacttgctcctgaatgaagactgtcctcagttcacaaccttcgcagccacaa 1151
Db 418 CACCCAGCAGCTTGGTCTCTGAATGAGAACTGTGCTCAGTTACAA-CTTCGGCAGCCACAA 476

QY 1152 gattcaaacccagctcaacctcatccacccggacatcttccccctgctcaccagtttccg 1211
Db 477 GATTCAAAACCCAGCTCAACCTCATCCA-CCGGACATCTTCCCTGTGCTCACCAGTTTCCG 535

QY 1212 ctgtaagaagaggggccccacacctcagtggtcccatggtccaggtgaaatgcctcccaa 1271
Db 536 CTGTAAGAAGAGAGGG-CCCAACCTCAGTGTGCCCATGTTTCAAGGTGAATGCTCCTCAA 594

QY 1272 gtaccagctccgtcccagagagagtgccagaggaatgcc-----attattactgca 1324
Db 595 GTACCAGTCCGTCCCAGAGAGGAGTGGCAGAGGGATGCCCATTTATATACTTTGACAAATC 654

QY 1325 atcctgaggaattcatagttgaggcgctcagcttcccacacttcca-gcagagcgtgcag 1383
Db 655 CTGTCACGGGAATCCATAGATAGGCGCGAGAGATACCAAGATACGAGCAGGAGCGGCAG 714

QY 1384 gagtacagagagagtgctgcagcagcggccccagccccagagagaaagaagtcagttacca 1443
Db 715 GAGTCCAGGAGGAAGTGCACAGAGGGGCCAG-CCCAGAGCGGAAAGAGAGTGAAGACCCA 773

QY 1444 gaatacctctccttggaacaggggtctgcc 1473
Db 774 GAATACAGCTCAGGAAAGGGCGCAACC 803

Search completed: February 18, 2001, 07:27:13
Job time: 22508 sec

251	ArgGlyAsnPheLeuValLeuLysAlaLysGluMetGlyLeuProValGl	267
252		
253		
254		
825	AGAGAAACTCTTGGTGCTCAAGACCAAGAGAGATGGCCTCCCAAGTTGG	874
267	yThrAlaAlaIleAlaProfileIleAlaAlaValLysAspGlyLysSerI	284
268		
269		
875	GACAGCTGCCATCGTCCCATCATGTGCTGTCAAGACGCGGAAAGCA	924
284	leThrHisGluGlyArgGluIleLeuAlaGluGluLeuCysThrProPro	300
285		
286		
925	TCACCTCATGAAGGAAGAGATTTGGCTGAAGAGCTGTGTACTCTCCCA	974
301	AspProGlyAlaAlaPheValValGluCysProAspGluSerPheIl	317
302		
975	CATCTGGTGCTGCTTTGGTGCTGGTGAATGTCAGATGAAGACTTCAT	1024
317	egInProIleCysGluAsnAlaThrPheGlnArgTyrGlnGlyLysAlaA	334
318		
1025	TCAACCCATCTGTGAGAATGCCACCTTTCAGAGGTACCAAGGAAGGCAG	1074
334	spAlaProValAlaLeuValValHisMetAlaProAlaSerValLeuVal	350
335		
1075	ATGCCCCCGTGGCCTTGGTGCTTCCATATGGCCCCAGACATCTGTCTGTG	1124
351	AspSerArgTyrGlnGlnTrpMetGluArPheGlyProAspThrGlnHi	367
352		
1125	GACAGCAGGTACCAAGCAGTGTGATGGAGAGTTTGGGCTTGACACCCAGCA	1174
367	sLeuValLeuAsnGluAsnCysAlaSerValHisAsnLeuArgSerHisL	384
368		
1175	CTTGGTCTGAATGAGAATCTGTGCCTCAGTTTCAACACTTCGCAGCCACA	1224
384	ysIleGlnThrGlnLeuAsnLeuIleHisProAspIlePheProLeuLeu	400
385		
1225	AGATTCAAAACCCAGCTCAACCTCATCCACCCGGACATCTTCCCCCTGCTC	1274
401	ThrSerPheArgCysLysLysGluGlyProThrLeuSerValProMetVa	417
402		
1275	ACCAAGTTTCCGCTGT...AGGAGGGCCCCACCCCTCAGTGTGCCATGGT	1321
417	lGlnGlyGluCysLeuLeuLysTyrGlnLeuArgProArGargluTrpG	434
418		
1322	TCAGSGTGAATGCCCTCTCCTCAAGTACCAGCTCCGTCACAGGAGGAGTGC	1371
434	lnArgAspAlaIleIleThrCysasnProGluGluPheIleValGluAla	450
435		
1372	AGAGGGATGGCATATTACTTGTCAATCCTCAGGAATTCATAGTTAGCGCG	1421
451	LeuGlnLeuProAsnPheGlnGlnSerValClnGluTrpArgArgSerAl	467
452		
1422	CTGCAGCTTCCCAACTTCCACAGAGGGCTGCAGAGGTACAGGAGAGTGC	1471
467	agInAspGlyProAlaProAlaGluLysArgSerGlnTyrProGluIleI	484
468		
1472	GCAGAGCGGCCACGCCCCAGCAGAGAAAGAAGTCAGTACCAGAAATCA	1521
484	lePheLeuGlyThrGlySerAlaIleProMetLysIleArgAsnValSer	500
485		
1522	TCCTTCCTGGAAACAGGGTCTGCCATCCCGATGAAGATTCGAAATGTCA	1571
501	AlaThrLeuValAsnIleSerProAspThrSerLeuLeuLeuAspCysGl	517
502		
1572	GCCACACTTGTCAACATAAGCCCGACACGCTCTGTGCTACTGGACTGTGG	1621
517	yGluGlyThrPheGlyGlnLeuCysArgHisTyrGlyAspGlnValAspA	534
518		
1622	TGAGGGCACATTTGGGCAGCTGTGCCGCTCATTACGGAGACACAGGTGGACA	1671
534	rgValLeuGlyThrLeuAlaAlaValPheValSerHisLeuHisAlaAsp	550
535		
1672	GGGTCTCTGGGCACCCCTGGCTGCTGTGTTGTGTCCTCCACCTGCACGACAGAT	1721

614 AlaGluIleSerSerPro.....AlaValGluArgLeuIleSe 626
: : : : :
1834 GAACCTCTTGGCTAGTCCCTCGAGCGGTGAACAAGTTGAACGCTA..... 1878
:
626 rSerLeuLeuArgThrCysAspLeuGluGluPheGlnThrCysLeuValA 643
: : : : :
1879GGAAATACGTCCCATATCCACCTGCTAGTTA 1909
:
643 rHisCysLysHisAlaPheGlyCysAlaLeuVal..... 654
: : : : :
1910 GGCACGTGCCCAACTCTTTTCGGAATAAGCCTAACTCTGGCGGCAAGCAC 1959
:
655 HisThrSerGlyTrpLysValValTyrSerGlyAspThrMetProCysG 671
: : : : :
1960 AATAGCGAACCCTGCTCAAGATCACTTACACGGCGACACTATGCATGTCA 2009
:
671 uAlaLeuValArgMetGlyLysAspAlaThrLeuLeuIleHisGluAla 688
: : : : :
2010 GGATCTGATCGATCTGGCGGCGGACTCCACCGTCTTATTCAGAGCGGA 2059
:
688 hrLeuGluAspGlyLeuGluGluAlaValGluLysThrHisSerThr 704
: : : : :
2060 CAATGGAGGAGCTTGGAGGAGGCGCGGTTAAAGAGCGCACACACC 2109
:
705 ThrSerGlnAlaIleSerValGlyMetArgMetAsnAlaGluPheIleMe 721
: : : : :
2110 GTGTCGACGGCTATTTCAGCAGGGCGGAATATGACGCCGCCATACGAT 2159
:
721 tIleuAsnHisPheSerGlnArgTyrAlaLysValPro.....LeuPheSerP 737
: : : : :
2160 TCTACCCCACTTTTCGACGGTTACGCTAAGTCCCGCGGATTAACCGAGTG 2209
:
737 roAsnPheSerGluLysValGlyValAlaPheAspHisMetLysValCys 753
: : : : :
2210 ATGAGGATATGCAACGAGTTCGAATAGCTTTGATATATGGAAGTGACC 2259
:
754 PheGlyAspPheProThrMetProLysLeuIleProProLeuLysAlaLe 770
: : : : :
2260 ATTGAGGATCTGCAGCACTACCAACAGCTCTACCCGCCCTTTTCGCGAT 2309
:
770 uPheAlaGlyAspIleGluGluMetGluGluArgArgGluLysArgGluL 787
: : : : :
2310 GTACGCCGAGTACACGGAAGACTGAGGAGCGGGCTGTCAAGCGGAGC 2359
:
787 euArgGlnValArgAlaAlaLeuLeuSerArgGluLeuAla 800
: : : : :
2360 TGAACACAGGCGCA.....AAGCGAAACTGGCA 2388
:
seq_name: gb_htgl.AC007352

seq_documentation_block:
LOCUS AC007352 190574 bp DNA HTG 24-FEB-2000
DEFINITION Drosophila melanogaster chromosome 2 clone BACR19J16 (D626) RPCI-98
19.J.16 map 47A-47B strain Y; cn bw sp, *** SEQUENCING IN PROGRESS
***, 36 unordered pieces.
ACCESSION AC007352
VERSION AC007352.8 GI:7025687
KEYWORDS HTG; HTGS_PHASE1.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
AUTHORS
Celisner, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G.,
Butenihoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,
Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,
Hinkle, A., Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K.,
Kearney, L., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P.,
Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S.,
Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Snir, E.,
Richards, S., Sethi, H., Svirkas, R.R., Wan, K.H., Webster, D.,
Woolley, P., Yang, S., Yee, M., Yu, C. and Rubin, G.M.

Sequencing of Drosophila melanogaster

UNPUBLISHED
JOURNAL
REFERENCE
AUTHORS

2 (bases 1 to 190574)
Celisner, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G.,
Butenihoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,
Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,
Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L.,
Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P.,
Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S.,
Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Snir, E.,
Svirkas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and
Rubin, G.M.

Direct Submission

UNPUBLISHED
JOURNAL
COMMENT

Submitted (22-APR-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
On Feb 24, 2000 this sequence version replaced gi:7018750.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (<http://www.fruitfly.org/sequence/>) or send email
to bdgp@fruitfly.berkeley.edu. All contigs in this submission meet
the following cutoffs: length >= 200 bases.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 36 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 678: contig of 678 bp in length
* 679 758: gap of unknown length
* 759 1358: contig of 600 bp in length
* 1359 1438: gap of unknown length
* 1439 2373: contig of 935 bp in length
* 2374 2453: gap of unknown length
* 2454 2962: contig of 509 bp in length
* 2963 3042: gap of unknown length
* 3043 3889: contig of 847 bp in length
* 3890 3970: contig of 1468 bp in length
* 3971 5517: gap of unknown length
* 5518 6572: contig of 1055 bp in length
* 6573 8652: gap of unknown length
* 8653 8387: contig of 1735 bp in length
* 8388 8467: gap of unknown length
* 8468 9060: contig of 593 bp in length
* 9061 9140: gap of unknown length
* 9141 11167: contig of 2027 bp in length
* 11168 11247: gap of unknown length
* 11248 14518: contig of 3271 bp in length
* 14519 14598: gap of unknown length
* 14599 18917: contig of 4239 bp in length
* 18918 24324: contig of 5407 bp in length
* 24325 24405: gap of unknown length
* 24406 29984: contig of 5580 bp in length
* 29985 30064: gap of unknown length
* 30065 35794: contig of 5730 bp in length
* 35795 35874: gap of unknown length
* 35875 44808: contig of 8934 bp in length
* 44809 44888: gap of unknown length
* 44889 58222: contig of 13334 bp in length
* 58223 58302: gap of unknown length
* 58303 78322: contig of 20020 bp in length
* 78323 78402: gap of unknown length
* 78403 122449: contig of 43847 bp in length
* 122450 123259: gap of unknown length
* 123260 179102: contig of 56773 bp in length
* 179103 179182: gap of unknown length
* 179183 179407: contig of 225 bp in length
* 179408 179487: gap of unknown length
* 179488 180592: contig of 1105 bp in length
* 180593 180672: gap of unknown length

```

* 180673 181383: contig of 711 bp in length
* 181384 181463: gap of unknown length
* 181464 182092: contig of 629 bp in length
* 182093 182172: gap of unknown length
* 182173 183480: contig of 1308 bp in length
* 183481 183560: gap of unknown length
* 183561 183771: contig of 211 bp in length
* 183772 183851: gap of unknown length
* 183852 184058: contig of 207 bp in length
* 184059 184138: gap of unknown length
* 184139 185035: contig of 897 bp in length
* 185036 185115: gap of unknown length
* 185116 185846: contig of 731 bp in length
* 185847 185926: gap of unknown length
* 185927 186560: contig of 634 bp in length
* 186561 186640: gap of unknown length
* 186641 187253: contig of 613 bp in length
* 187254 187333: gap of unknown length
* 187334 187611: contig of 278 bp in length
* 187612 187691: gap of unknown length
* 187692 188409: contig of 718 bp in length
* 188410 188489: gap of unknown length
* 188490 189196: contig of 707 bp in length
* 189197 189276: gap of unknown length
* 189277 189969: contig of 693 bp in length
* 189970 190049: gap of unknown length
* 190050 190574: contig of 525 bp in length.
FEATURES
    Location/Qualifiers
        1..190574
            /organism="Drosophila melanogaster"
            /strain="y; cn bw sp"
            /db_xref="taxon:7227"
            /chromosome="2"
            /map="47A-47B"
            /clone="BACR19J16 (D626) RPCI-98 19.J.16"
            /clone_lib="RPCI-98 (Roswell Park Cancer Institute
Drosophila melanogaster BAC library, partial EcoRI in
pBAC3.6"
BASE COUNT 53370 a 39726 c 40358 g 54299 t 2821 others
ORIGIN

alignment_scores:
    Quality: 1136.00      Length: 862
    Ratio: 2.197          Gaps: 31
    Percent Similarity: 59.977 Percent Identity: 34.339

alignment_block:
US-09-434-382-2 x AC007352
Align seg 1/1 to: AC007352 from: 1 to: 190574
9 ArgSerAlaAlaGlyArgThrMetSerGlnGlyArgThrIleSerGlnAl 25
||||| ||||| : : : : : ||||| |||||
131416 AGTTCGCGGATCTACAGACCTTG...CGACGGCTCAACACGATTAATT 131462
||| ||||| : : : : : ||||| |||||
25 aProAlaArgArgGluArgProArgLysAspProLeuArgHisLeuArgt 42
||| ||||| : : : : : ||||| |||||
131463 GATGCGGCCCAATTCCTCGGCCAAGCATCTTTA.....ACAGGAC 131506
||| ||||| : : : : : ||||| |||||
42 hrArgGluLysArgGlyProSerGly.....CysSerGly 53
||| ||||| : : : : : ||||| |||||
131507 CTCGCTATGACGGGAACCCCAATGTGCTAGAAAAAAGCTGCCCTCGGTG 131556
||| ||||| : : : : : ||||| |||||
54 GlyProAsnThrValTyrLeuGlnValValAlaAlaGlySerArgAspSe 70
||||| ||||| : : : : : ||||| |||||
131557 GTGCGGGAACAGTAATCTTCAGGTCTCGGCTCTGGCGCAATGGAGC 131606
||||| ||||| : : : : : ||||| |||||
70 rGlyAlaAlaLeuTyrValSerGlnPheAsnArgTyrIlePheAsnC 87
||||| ||||| : : : : : ||||| |||||
131607 GCCTGCCGCTGTTATTGTTTACGACCAAGCAGCGATCTCTTCAACT 131656
||||| ||||| : : : : : ||||| |||||
87 ySgLyGluGlyValGlnArgLeuMetGlnGluHisLysLeuLysValAla 103
||||| ||||| : : : : : ||||| |||||

```

```

||||| ||||| : : : : : ||||| |||||
131657 GCGGAGGGAACCCAGAGACTAGCTCAGAGCACAACACCGCCTATCA 131706
||||| ||||| : : : : : ||||| |||||
104 ArgLeuAspAsnIlePheLeuThrArgMetHisTrpSerAsnValGlyG 120
||||| ||||| : : : : : ||||| |||||
131707 CCCTGGGACAAATTTCTTACACAAATACGTGGGCTTCTCTGGGAGG 131756
||||| ||||| : : : : : ||||| |||||
120 yLeuSerGlyMetIleLeuThrLeuLysGluThrGlyLeuProLysCysv 137
||||| ||||| : : : : : ||||| |||||
131757 GTTACCAGGTTTGACGCTCACCATCCAGATCGCGGAGTGCAGCATTTG 131806
||||| ||||| : : : : : ||||| |||||
137 alLeuSerGlyProProGlnLeuGluLysTyrLeuGluAlaIleLysle 153
||||| ||||| : : : : : ||||| |||||
131807 GACTCCATGGACCGCTCATCTCGGCTCTATGCTGCAGTCAATGGACGT 131856
||||| ||||| : : : : : ||||| |||||
154 pheSerGlyProLeuLysGlyIleGluLeuAlaValArgProHisSerAl 170
||||| ||||| : : : : : ||||| |||||
131857 TTCGTG...GTGCTAAAGAAATCTCCAG.....GTGCGGCCAAGCATTTG 131897
||||| ||||| : : : : : ||||| |||||
170 aProGlu.....TyrGluAspGluThrMetThrValTyrGlnIleP 184
: ||| : : : : : ||||| |||||
131898 TTCTCAGGGCGGCTCTTTTGGAGATTCCATCTTGAAGTGGACTCTCTGC 131947
||||| ||||| : : : : : ||||| |||||
184 ro...IleHisSerGluGlnArgArgGlyLysHisGlnProTrpGlnSer 199
||| ||||| : : : : : ||||| |||||
131948 CCTGATCAATTCCGAA..... 131964
||||| ||||| : : : : : ||||| |||||
200 ProGluArgProLeuSerArgLeuSerProGluArgSerSerAspSerGl 216
: ||| : : : : : ||||| |||||
131965 .....GATCCGACGAAAAAGT..... 131979
||||| ||||| : : : : : ||||| |||||
216 uSerAsnGluAsnGluProHisLeuProHisGlyValSerGlnArgArg 233
||||| ||||| : : : : : ||||| |||||
131979 ..... 131979
233 lyValArgAspSerSerLeuValValAlaPheIleCysLysLeuHisLeu 249
||||| ||||| : : : : : ||||| |||||
131980 .....GTAATTAACATACATTTGTACAGCTGAAACCG 132009
||||| ||||| : : : : : ||||| |||||
250 LysArgGlyAsnPheLeuValLeuLysAlaLysGluMetGlyLeuProVa 266
: ||| : : : : : ||||| |||||
132010 CGAGCTGGCGGCTTAACCTTGGTGAAGTGGTGGAGCAAGGTGTCGCCG 132059
||||| ||||| : : : : : ||||| |||||
132060 AGGA.....CCACTTCTGGGACAGCTAAAGAATGCCAACG 132094
||||| ||||| : : : : : ||||| |||||
283 erIleThr...HisGluArgGluIleLeuAlaGluGluLeuLeuCysThr 298
: ||| : : : : : ||||| |||||
132095 ATATCACACTGCCTGATGGAAGTGGTGGCTTTCAGTCGACGTCACCGAA 132144
||||| ||||| : : : : : ||||| |||||
299 ProProAspProGlyAlaAlaPheValValGluCysProAspGlus 315
: ||| : : : : : ||||| |||||
132145 GCCAGTGAACACTGCATCTCTTGTTCCTTGTATGTGCCAGTGAGAA 132194
||||| ||||| : : : : : ||||| |||||
315 rPheIleGlnProIleCysGluAsnAlaThrPheGlnArgTyrGlnGlyL 332
: ||| : : : : : ||||| |||||
132195 CTATCTACCAGCACTGTTACACACGCAAAAGGCTTAAAGATTAGCGG 132244
||||| ||||| : : : : : ||||| |||||
332 ysAlaAsp...AlaProValAlaLeuValValHisMetAlaProAlaSer 347
: ||| : : : : : ||||| |||||
132245 AAGAGAAGTTAACTGAAGTGGCTTGTGTGTCACCTTCACGCTTATCAT 132294
||||| ||||| : : : : : ||||| |||||
348 ValLeuValAspSerArgTyrGlnGlnTrp...MetGluArgPheGlyPr 363
: ||| : : : : : ||||| |||||
132295 ATATCTTCTCGTCAGGTGTACAGGATTTTGTGGTGGAGAGTTTTCATC 132344
||||| ||||| : : : : : ||||| |||||
363 oAspThrGlnHisLeuValLeuAsnGlu.....AsnCysAlaSerValH 378
: ||| : : : : : ||||| |||||
132345 TGAGSCACAGCACATATACTTAAGTTCCTCCCTCAACCAATTCAGGCT 132394
||||| ||||| : : : : : ||||| |||||
378 isAsnLeuArgSerHisLysIleGlnTrpGlnLeuAsnLeuIleHisPro 394
: ||| : : : : : ||||| |||||

```


[illegible]

seq_documentation_block:
LOCUS AEO03830 261846 bp DNA INV 04-OCT-2000
DEFINITION *Drosophila melanogaster* genomic scaffold 142000013386047 section 13
of 52, complete sequence.

01 327 complete sequence.
AE003830 AE002787
AE003830.2 GI:110727666
HTG.

fruit fly,
Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 261846)

Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
Ananastadis,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
Brandon,R.C., Rogers,Y.H., Blazer,R.G., Champe,M., Pfeiffer,B.D.,
Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor
Miklos,G.L., Abrij,J.F., Agbayani,A., An,H.J.,
Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Basu,A.,
Baxendale,J., Bayraktarolu,L., Beasley,E.M., Beeson,K.Y.,
Benos,P.V., Bertram,B.P., Bhandari,D., Bolshakov,S., Borokova,D.,
Botchan,M.R., Bouck,J.J., Brokstein,P., Brottier,P., Burtis,K.C.,
Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferreira,S.,
Fleischmann,W., Fosler,C., Gabrielian,A.E., Garg,N.S.,

REFERENCE

AUTHORS

ORGANISM

SOURCE

KEYWORDS

VERSION

ACCESSION

Gelbart, W.M., Glasser, K., Glodek, A., Gong, F., Gorrell, J.H., Gu, Z., Guan, P., Harris, M., Harris, N.L., Harvey, D., Heiman, I.J., Hernandez, J.R., Houck, J., Hostin, D., Houston, K.A., Howland, T.J., Wei, M.H., Ibegwam, C., Jatali, M., Kalush, F., Karpen, G.H., Ke, Z., Kennison, J.A., Ketchum, K.A., Kimmel, B.E., Kodira, C.D., Kraft, C., Kravitz, S., Kulp, D., Lai, Z., Lasko, P., Lei, Y., Levitsky, A.A., Li, J., Li, Z., Liang, Y., Lin, X., Liu, X., Mattei, B., McIntosh, T.C., McLeod, M.P., McPherson, D., Merkulov, G., Milshina, N.V., Mobarri, C., Morris, J., Moshrefi, A., Mount, S.M., Moy, M., Murphy, B., Murphy, L., Muzny, D.M., Nelson, D.L., Nelson, D.R., Nelson, K.A., Nixon, K., Nusskern, D.R., Pacleb, J.M., Palazzolo, M., Pittman, G.S., Pan, S., Pollard, J., Puri, V., Reese, M.G., Reinert, K., Remington, K., Saunders, R.D., Scheeler, F., Shen, H., Shue, B.C., Siden-Kiamos, I., Simpson, M., Skupski, M.P., Smith, T., Spier, E., Spradling, A.C., Stapleton, M., Strong, R., Sun, E., Swirskas, R., Tector, C., Turner, R., Venter, E., Wang, A.H., Wang, X., Wang, Z.Y., Wassarman, D.A., Weinstein, G.M., Weissbach, J., Williams, S.M., Woodage, T., Worley, K.C., Wu, D., Yang, S., Yao, Q.A., Ye, J., Yeh, R.F., Zaveri, J.S., Zhan, M., Zhang, G., Zhao, Q., Zheng, L., Zheng, X.H., Zhong, F.N., Zhong, W., Zhou, X., Zhu, S., Zhu, X., Smith, H.O., Gibbs, R.A., Myers, E.W., Rubin, G.M. and Venter, J.C.

The genome sequence of *Drosophila melanogaster*
Science 287 (5461), 2185-2195 (2000)
20196006
2 (bases 1 to 261846)
Adams, M.D., Celniker, S.E., Gibbs, R.A., Rubin, G.M. and Venter, C.J.
Direct Submission
Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
On Oct 9, 2000 this sequence version replaced gi:7303755.
Location/Qualifiers
1. 261846
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/chromosome="2R"
join(960..1069,1135..1237,1290..1357,1416..1731,
2658..2759,3801..4142,4200..4354,4727..4835,6101..6289,
6499..6941,6999..7841,7921..8315)
/gene="Pfk"
/product="Ct13302"
/db_xref="FLYBASE:FBan0004001"
/db_xref="FLYBASE:FBgn0003071"
<960..>8315
/gene="Pfk"
/note="CG4001"
/map="46E3-46E6"
/db_xref="FLYBASE:FBan0004001"
/db_xref="FLYBASE:FBgn0003071"
join(1026..1069,1135..1237,1290..1357,1416..1731,
2658..2759,3801..4142,4200..4354,4727..4835,6101..6289,
6499..6941,6999..7841,7921..8059)
/gene="Pfk"
/note="Pfk gene product [alt 1]"
/codon_start=1
/db_xref="FLYBASE:FBan0004001"
/db_xref="FLYBASE:FBgn0003071"
/protein_id="AAF58840.1"
/db_xref="GI:7303792"
/translation="MHSIKFRVFTKLKPIFLINGRIPICRHPGPTTFRLEISNKT
PIROKTFNIGTCVTRSHLCCPRDISGNLTLLSVKFNCKRHCICKLSDSGDKNDS
GKNIQDKSAOCPKPIINLNGFLNAYSEKNVKKKSPKPKCKSVDELKRC
LRTMQVDIVHPFKKGLKGLAVFTSGDSDGQMAAVRACVRMAIYLGCXVYFIREG
YQGMVDGDCIOEAWNVSSIIHRGTTIGSAGDQFRFCRQLKAANLQRTGTLN
LVVIGDGLTALHRIEADAISSTAYSHORTFIMEVGRHCYLAIVLGVGLACEDFI
FTPEMPKVDWDRCSQLOAORSAGORLNIVIVAEAGMDRECHPITAEVKKVIDER
LKHDARITVLGHVORGNPSAFDRILACRMGAETLALMEATKDSVPVYSLDGNV
RPLMECVERTQAVAKMAEKRWADAVKLRGSRFERNLETIKMLTRKLPKPKENADG
KLEGYRLAVMHIGAPACMAAARFVRNAIYRGDVYINDVGEGLTAGNVRELWG
SDVSGVGGGAYLGRKTLPEGKFEIAARKEFKIQLGLIIGFSFYHAAGQIADQ
IETMPPQIPVIVIPSTISNNVPTFSLGCDTGLNEITEICDRIRQSAQGTKRFRVF
IETMPPQICGLATLAGAGDAAYIYEKFSIKDLOQDVHMAKMGVSRGLTLR
NEKASENSTDYFIYRLYSEEGKGLFCRNLGHMQGGSPTFFDRNMGTKRAKACVD

WLAQAQKANTDANGVNVCKSPDTATLLGIVSRQYRFSPLVDLIAETNDFORIPKKQMW
LRLPRLILAKHDSAYEEGMYITVEECDTDAV"
join(2152..2759,3801..4142,4200..4354,4727..4835,
6101..6289,6499..6941,6999..7841,7921..8315)
/gene="Pfk"
/product="Ct13275"
/db_xref="FLYBASE:FBan0004001"
/db_xref="FLYBASE:FBgn0003071"
join(2613..2759,3801..4142,4200..4354,4727..4835,
6101..6289,6499..6941,6999..7841,7921..8059)
/gene="Pfk"
/note="Pfk gene product [alt 2]"
/codon_start=1
/db_xref="FLYBASE:FBan0004001"
/db_xref="FLYBASE:FBgn0003071"
/protein_id="AAF58841.1"
/db_xref="GI:7303793"
/translation="MNSEINQRELARGSQDKGLAVTSGDSDGQMAAARVACVRMAI
YLGCXVYFIREGQYGVGGDCIOEAWNVSSIIHRGTTIGSAGDQFRFCRQLK
AANLQRTGTLNIVIGDGLTALHRIEADAISSTAYSHORTFIMEVGRHCYLA
VLGVSTINDFCGTDMTIGTDTALHRIEADAISSTAYSHORTFIMEVGRHCYLA
LVGLACEDFIPEMPKVDWDRCSQLOAORSAGORLNIVIVAEAGMDRECHP
TAEVKKVIDERLKHDAITVLGHVORGNPSAFDRILACRMGAETLALMEATKDSV
PVYSLDGNVAVRPLMECVERTQAVAKMAEKRWADAVKLRGSRFERNLETIKMLTR
LKPKPKENADGKLEGYRLAVMHIGAPACMAAARFVRNAIYRGDVYINDVGE
GLTAGNVRELGSVSGVGGGAYLGRKTLPEGKFEIAARKEFKIQLGLIIGF
ESYHAAGQIADQNDYPOFCPIVIVIPSTISNNVPTFSLGCDTGLNEITEICDR
QSAQGTKRFRVFITMGGYCGYLAIVLGVGLACEDFIYEKFSIKDLOQDVHMAK
MAREVSRGLTLRNEKASENSTDYFIYRLYSEEGKGLFCRNLGHMQGGSPTFFDR
NMGTKRAKACVDWLAQAQKANTDANGVNVCKSPDTATLLGIVSRQYRFSPLVDLIA
ETNDFORIPKKQMWLRLPRLILAKHDSAYEEGMYITVEECDTDAV"
join(8915..9052,9216..9726,10663..10911,10976..11174,
11235..11542,11596..>12210)
/gene="CG11866"
/product="Ct37014"
/db_xref="FLYBASE:FBan0011866"
/db_xref="FLYBASE:FBgn0033486"
<8915..>12210
/gene="CG11866"
/map="46B4-46E6"
/db_xref="FLYBASE:FBan0011866"
/db_xref="FLYBASE:FBgn0033486"
join(9286..9726,10663..10911,10976..11174,11235..11542,
11596..12210)
/gene="CG11866"
/note="CG11866 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0011866"
/db_xref="FLYBASE:FBgn0033486"
/protein_id="AAF58839.1"
/db_xref="GI:7303791"
/translation="MVSTQMCSGGTSASEGVSPTFGAGSGSASCTGMASGAVSSG
SGTVQPEPPESTANRSSVLRRSTSYOGNOAHSSYRVVSAGSLAESASVTRLSHYTHQ
MQQRPTATTTTTSAPSLYDLNELLIEKILSYVDYKKVSNLRLVSHRNDICMAML
NATFKQIKTTLRSFOAKMSPRRSHRRNHPLACEDIIETCYRLSLQLQSMGKH
IERGHCCFFGAILDEVQAILNYSITPRLQRYRYTDELDTLWAMEYFKDRIEAT
LPGYAKNDFTYLTPTTKRPTLAISDLESASSNPQNMVLRKIRKIKQGMKY
NNOLSVYRLTRTCRKAQSKOLAESNLISEQOKQTLVYANRLDENDKKNEMS
KFTLLQELNCKKTELQFWRSKSPAIYAVCSNOKVAVVPPEDQALVQGVDPFH
FIINEDVAESDVSGELKEFASPESTAKLAVSTAKNLKRMKPSSESQSNATAS
TSKAAVSQSQPMATGAAKAGGYSPLKFTGNHQRSGVIVSPVSNLGVAAAT
DNVPSKALEEPEEKARRVQKANNRYNTPKRSK"
complement(join(<12115..12937,12999..13642,13694..13708))
/gene="CG17753"
/product="Ct11457"
/db_xref="FLYBASE:FBan0017753"
/db_xref="FLYBASE:FBgn0033487"
complement(<12115..>13708)
/gene="CG17753"
/map="46B4-46E6"
/db_xref="FLYBASE:FBan0017753"
/db_xref="FLYBASE:FBgn0033487"
complement(join(12802..12937,12999..13642,13694..13708))

```

/gene="CGI7753"
/note="CGI7753 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0017753"
/db_xref="FLYBASE:FBgn0033487"
/protein_id="AAF58838.1"
/db_xref="GI:7303790"
/translation="MSSIKIEFAVMRRGDSYAGALRSALDGVGVDEIDTQEGRVII
QTPRWSEIQDKIEATGVRVILSGFGQSAGVALINTGVSVDKTPIQGVVRFITAD
KKPGVVDVGVDSGLHGLHIESGDSVGEHYNPRGSPHGPSAAGAEERH
AGDLGNTRADENGRAFRFVDPVLEVMVDIIGRAVVLTAADDLGRGNDQSLDNGSG
ERIAACGIIARSAGILENFKKRICACDGVTLWDERNKPLAGKERSOKL"
join(<18636..18761,19765..19876,19968..20213,22806..22878,
22958..23344,23489..23633,23934..24407,25026..25202)
/gene="CGI1867"
/product="CT37016"
/db_xref="FLYBASE:FBan0011867"
/db_xref="FLYBASE:FBgn0033488"
<18636..25202
/gene="CGI1867"
/map="46E4-46E8"
/db_xref="FLYBASE:FBan0011867"
/db_xref="FLYBASE:FBgn0033488"
join(18636..18761,19765..19876,19968..20213,22806..22878,
22958..23344,23489..23633,23934..24407,25026..25202)

alignment_scores:
  Quality: 1136.00      Length: 862
  Ratio: 2.197          Gaps: 31
Percent Similarity: 59.97 Percent Identity: 34.33

alignment_block:
US-09-434-382-2 x AE003830/rev ..
Align seg 1/1 to reverse of: AE003830 from: 1 to: 261846

9 ArgSerAlaAlaClyArgThrMetSerGlnGlyArgThrIleSerGlnAl 25
|||||
200687 AGGTCGCCGATCTACAGAACCTTG...CGAACGCTCACACACCACTAATTT 200641
|||
25 aProAlaArgGluArgProArgLysAspProLeuArgHisLeuArgT 42
|||
200640 GATGGCGGCCACAAATTCCTCGCCCAAGGATCCTTTA.....ACAGGAC 200597
42 hrArgGluLysArgGlyProSerGly.....CysSerGly 53
|||
200596 CTCGCTATGAGCGGGAACCAATGCTTAGAAAAAGCTGGCCTCGGTG 200547
54 GlyProAsnThrValTyrLeuGlnValAlaAlaGlySerArgaspse 70
|||||
200546 GTGCCGGGAACAGTAATCTCAGGTCTCTGGGCTCTGGCGCCCAATGGAGC 200497
70 rGlyAlaAlaLeuTyrValPheSerGluPheAsnArgTyrLeuPheAsnC 87
|||||
200496 GCCTCGCGCTGTTATTGTTAGGACGACGAGCAGGACGCTATCTCTCAACT 200447
87 yGlyGluGlyValGlnArgLeuMetGlnGlnHisLysLeuLysValAla 103
|||||
200446 CGCGCGAGGGAACCCAGAGACTAGCTACGAGCAGACAAACCCGCCCTATCA 200397
104 ArgLeuAspAsnIlePheLeuThrArgMetHisTrpSerAsnValGlyGI 120
|||||
200396 CGCCTGGAAACAAATTTTCCTTACACAAAATACGTGGGCTCTCTGCGGAGG 200347
120 yLeuSerGlyMetIleLeuThrLeuLysGluThrGlyLeuProLysCysV 137
|||||
200346 GTTACCAGGTTTGACCTCACCATCCAAAGATCCGAGTGCGGCACATTG 200297
137 alLeuSerGlyProProGlnLeuLysTyrLeuGluAlaIleLysIle 153
|||||
200296 GACTCCATGACCGCCTCATCTCGGCTCTATGCTGCAGTCAATGCCGACGT 200247
154 PheSerGlyProLeuLysGlyIleGluLeuAlaValArgProHisSerAl 170

```

```

200246 TTCGTG...GTGCTAAAGATCTCCAG.....GTGCGGCCAAACGATTG 200206
170 aProGlu.....TyrGluAspGluThrMetThrValTyrGlnIleP 184
|||
200205 TTCTGAGGGCCCTGTTTTCAGGATTCATCTTGAAGCTGGACTCTGCG 200156
184 ro...IleHisSerGluGlnArgGlyLysHisGlnProTrpGlnSer 199
|||
200155 CCCTGATCAATTCGAA..... 200139
200 ProGluArgProLeuSerArgLeuSerProGluArgSerSerAspSerGI 216
|||
200138 .....GATCCGACGAAAGT..... 200124
216 uSerAsnGluAsnGluProHisLeuProHisGlyValSerGlnArgArg 233
200124 ..... 200124
233 lyValArgAspSerSerLeuValAlaPheIleCysLysLeuHisLeu 249
|||
200123 .....GTAATTAACTACATTTGTCAGCTGAAACCG 200094
250 LysArgGlyAsnPheLeuValLeuLysAlaLysGluMetGlyLeuProVa 266
|||
200093 CGAGCTGGCGCCCTAAACTTGTGAAGTGGGTGGAGCAAGGTGTCGCC 200044
266 lGlyThrAlaAlaIleAlaProIleIleAlaValLysAspGlyLys 283
|||
200043 AGGA.....CCACTTCTGGGACACCTAAAGATGGCAACG 200009
283 erIleThr...HisGluGlyArgGluIleLeuAlaGluGluLeuCysThr 298
|||
200008 ATATCACACTGCCTGATGAAAGGTGCTGCTTCAGTCGACGTCACCGAA 199959
299 ProProAspProGlyAlaAlaPheValValGluCysProAspGluSe 315
|||
199958 GCCAGTGAACACTGCATGCTCTTGTCTTCTGATGTCGCCAGTGAGAA 199909
315 rPheIleGlnProIleCysGluAsnAlaThrPheGlnArgTyrGlnGlyL 332
|||
199908 CTATCTACACGACTGCTTACACACGCAAGCTTAAAGATTAGCGG 199859
332 ysAlaAsp...AlaProValAlaLeuValValHisMetAlaProAlaSer 347
|||
199858 AAGAGAAGTTAACTGAAGTGGCGCTTGTGTGCACTTCACGCCCTATCAT 199809
348 ValLeuValAspSerArgTyrGlnGlnTrp...MetGluArgPheGlyPr 363
|||
199808 ATATCTTCTGCTCAGGTGTACAAGGATTTGTGTGTGAGAGTTTTCATC 199759
363 oAspThrGlnHisLeuValLeuAsnGlu.....AsnCysAlaSerValH 378
|||||
199758 TGAGGCACAGCACAATATCTAAGTTCTCCCTCAACCAATTCACGCT 199709
378 isAsnLeuArgSerHisLysIleGlnThrGlnLeuAsnLeuIleHisPro 394
|||
199708 AT...GCAGCGGCCCATCGCATTCACACCACTCCACCACTGGCACCG 199662
395 AspIlePheProLeuLeu...ThrSerPheArgCysLysLysGluGlyPr 410
|||||
199661 CAAGTATTTCACCTCTCTGGCGCAACAATTTCTCTGTCGAGAGTCAG.... 199617
410 oThrLeuSerValProMetValGlnGlyGlu..... 420
199616 .ACTCTCAGTCTAAACCTTTAAAAAACAAGATTGGACAGCGCCGATTCCG 199568
420 ..... 420
199567 AGGACAAAGCAATTCGAAAGCTTAACGAGACAGACAGAACAGGGTGTGGTC 199518
421 CysLeuLeuLysTyrGlnLeuArgProArgGluTrpGlnArgAspAl 437
|||||

```


62379	GCTTAGAAAAAGCTGGCCTCCGTGGTGGCGGAACAGTAATACTTCAGG	62428
62	alValAlaAlaaglySerArgAspSerGlyAlaAlaLeuTyrValPheSer	78
62429	TTCGGGCTCTGGCCCAATGGACGCCCTGCCGCTGTTATTGTTTACG	62478
79	GluPheAsnArgTyrLeuPheAsnCysGlyGluGlyValGlnArgLeuMe	95
62479	GACGAGCACGCTATCTCTTCACTGGGGGAGGAAACCCAGAGACTAGC	62528
95	tGlnGluHisLysLeuLysValAlaArgLeuAspAsnIlePheLeuThrA	112
62529	TCACGAGCACAACCCGCCCTATCACGCCCTGGAAACAATTTCTCTTAC	62578
112	rgMetHisTrpSerAsnValGlyGlyLeuSerGlyMetIleLeuThrLeu	128
62579	AAATACGTGGCTCTCTGGCGAGGTTACCAAGTTTGACCCCTACCATC	62628
129	LysGluThrGlyLeuProLysCysValLeuSerGlyProProGlnLeuG	145
62629	CAAGATGCCGAGTGGCGGACATGGGACTCCATGGACGGCTCATCTCG	62678
145	uLysTyrLeuGluAlaIleLysIlePheSerGlyProLeuLysGlyIleG	162
62679	CTCTATGCTGCAGTCAATGCGACGTTTTCGTG...GTGCTAAGAATCTCC	62725
162	luLeuAlaValArgProHisSerAlaProGlu.....TyrGluAsp	175
62726	AG.....GTCCGGCCAAACGATTGTTCTGAGGGCGGCTGTTTGAGGAT	62769
176	GluThrMetThrValTyrGlnIlePro...IleHisSerGluGlnArgAr	191
62770	TCCATCTTGAAGGTGGACTCTCTGCCCTGATCAATTTCGGAA.....	62811
191	gGlyLysHisGlnProTrpGlnSerProGluArgProLeuSerArgLeuS	208
62812G	62812
208	erProGluArgSerSerAspSerGluSerAsnGluAsnGluProHisLeu	224
62813	ATCCGACGAAAAGT.....	62826
225	ProHisGlyValSerGlnArgGlyValArgAspSerSerLeuValva	241
62827GTAAT	62831
241	lAlaPheIleCysLysLeuHisLeuLysArgGlyAsnPheLeuValLeuL	258
62832	TAACCTACATATTGTACGTGAAACCGCAGCTGGCGCCCTAAACTTG	62881
258	ysAlaLysGluMetGlyLeuProValGlyThrAlaAlaIleAlaProIle	274
62882	AGTGGCTGCAGCAAGTGTGGCCGAGGA.....CCACTT	62916
275	IleAlaAlaValLysAspGlyLysSerIleThr...HisGluGlyArgG	290
62917	CTGGGACAGCTAAAGAATGGCAACGATATCATCGCTGATGGAAAGGT	62966
290	uIleLeuAlaGluGluLeuCysThrProProAspProGlyAlaAlaPheV	307
62967	GGTGGCTTCAGTCGACGTCACCGAAGCCAGTGAAACTGCACGTGCTTTG	63016
307	alValValGluCysProAspGluSerPheIleGlnProIleCysGluAsn	323
63017	TTTTCCTTGATGTGCCAGTGAAGAACTATCCACGACGACTGCTTACAC	63066
324	AlaThrPheGlnArgTyrGlnGlyLysAlaAsp...AlaProValAlaLe	339
63067	GGCAAAAGGCTTAAAAAGTTAGGCGAAGAGAAAGTTAACTGAAGTGGCGCT	63116
339	uValValHisMetAlaProAlaSerValLeuValAspSerArgTyrGlnG	356

[illegible]


```
* 18751 18830: gap of unknown length
* 18831 19376: contig of 546 bp in length
* 19377 19456: gap of unknown length
* 19457 20255: contig of 799 bp in length
* 20256 20335: gap of unknown length
* 20336 20957: contig of 622 bp in length
* 20958 21037: gap of unknown length
* 21038 22187: contig of 1150 bp in length
* 22188 22267: gap of unknown length
* 22268 22871: contig of 604 bp in length
* 22872 22951: gap of unknown length
* 22952 24044: contig of 1093 bp in length
* 24045 24124: gap of unknown length
* 24125 25189: contig of 1065 bp in length
* 25190 25269: gap of unknown length
* 25270 26217: contig of 948 bp in length
* 26218 26297: gap of unknown length
* 26298 27312: contig of 1015 bp in length
* 27313 27392: gap of unknown length
* 27393 28934: contig of 1542 bp in length
* 28935 29014: gap of unknown length
* 29015 29998: contig of 984 bp in length
* 29999 30078: gap of unknown length
* 30079 31456: contig of 1378 bp in length
* 31457 31536: gap of unknown length
* 31537 32422: contig of 1706 bp in length
* 32423 33322: gap of unknown length
* 33323 34819: contig of 1497 bp in length
* 34820 34899: gap of unknown length
* 34900 35765: contig of 866 bp in length
* 35766 35845: gap of unknown length
* 35846 37032: contig of 1187 bp in length
* 37033 37112: gap of unknown length
* 37113 38879: contig of 1767 bp in length
* 38880 38959: gap of unknown length
* 38960 39889: contig of 930 bp in length
* 39890 39969: gap of unknown length
* 39970 40967: contig of 998 bp in length
* 40968 41047: gap of unknown length
* 41048 42656: contig of 1609 bp in length
* 42657 42736: gap of unknown length
* 42737 44172: contig of 1436 bp in length
* 44173 44252: gap of unknown length
* 44253 45669: contig of 1417 bp in length
* 45670 45749: gap of unknown length
* 45750 47740: contig of 1991 bp in length
* 47741 47820: gap of unknown length
* 47821 48827: contig of 1007 bp in length
* 48828 48907: gap of unknown length
* 48908 49770: contig of 863 bp in length
* 49771 49850: gap of unknown length
* 49851 51951: contig of 2101 bp in length
* 51952 52031: gap of unknown length
* 52032 53654: contig of 1623 bp in length
* 53655 53734: gap of unknown length
* 53735 55071: contig of 1337 bp in length
* 55072 55151: gap of unknown length
* 55152 56770: contig of 1619 bp in length
* 56771 56850: gap of unknown length
* 56851 58743: contig of 1893 bp in length
* 58744 58823: gap of unknown length
* 58824 60291: contig of 1468 bp in length
* 60292 60371: gap of unknown length
* 60372 63759: contig of 3388 bp in length
* 63760 63839: gap of unknown length
* 63840 66476: contig of 2637 bp in length
* 66477 66556: gap of unknown length
* 66557 69321: contig of 2765 bp in length
* 69322 69401: gap of unknown length
* 69402 70806: contig of 1405 bp in length
* 70807 70886: gap of unknown length
* 70887 73742: contig of 2856 bp in length
* 73743 73822: gap of unknown length

* 73823 75075: contig of 1253 bp in length
* 75076 75155: gap of unknown length
* 75156 77979: contig of 2824 bp in length
* 77980 78059: gap of unknown length
* 78060 81375: contig of 3316 bp in length
* 81376 81455: gap of unknown length
* 81456 84452: contig of 2997 bp in length
* 84453 84532: gap of unknown length
* 84533 87596: contig of 3064 bp in length
* 87597 87676: gap of unknown length
* 87677 102529: contig of 14853 bp in length
* 102530 102609: gap of unknown length
* 102610 103128: contig of 519 bp in length
* 103129 103208: gap of unknown length
* 103209 103812: contig of 604 bp in length
* 103813 103892: gap of unknown length
* 103893 104112: contig of 220 bp in length
* 104113 104192: gap of unknown length
* 104193 104767: contig of 575 bp in length
* 104768 104847: gap of unknown length
* 104848 105570: contig of 723 bp in length
* 105571 105650: gap of unknown length
* 105651 106151: contig of 501 bp in length
* 106152 106231: gap of unknown length
* 106232 106941: contig of 710 bp in length
* 106942 107021: gap of unknown length
* 107022 107785: gap of unknown length
* 107786 108412: contig of 627 bp in length
* 108413 108492: gap of unknown length
* 108493 109249: contig of 757 bp in length
* 109250 109329: gap of unknown length
* 109330 110040: contig of 711 bp in length
* 110041 110120: gap of unknown length
* 110121 110727: contig of 607 bp in length
* 110728 110807: gap of unknown length
* 110808 111413: contig of 606 bp in length

alignment_scores:
    Quality: 1128.00      Length: 811
    Ratio: 2.274         Gaps: 28
    Percent Similarity: 61.159      Percent Identity: 34.649

alignment_block:
US-09-434-382-2 x AC007417/rev ..

Align seg 1/1 to reverse of: AC007417 from: 1 to: 125448

55 ProAsnThrValTyrLeuGlnValAlaAlaGlySerArgAspSerG1 71
   |||:::||||| |||:::|||||:::|||||:::|||||:::
84410 CCGGGAACAGTAAATCTTCAGGTTCTGGGCTCTGGCGCAATGGAGGCC 84361

71 yAlaAlaLeuTyrValPheSerGluPheAsnArgTyrLeuPheAsnCysG 88
   |||:::|||||:::|||||:::|||||:::|||||:::|||||
84360 TGGCGCTGTTTATTGTTTACGGACACGACGCTATCTCTCAACTGCG 84311

88 lYGlUgLyValGlnArgLeuMetGlnGluHisLysLeuLysValAlaArg 104
   |||:::|||||:::|||||:::|||||:::|||||:::|||||
84310 GCAGGGAACCCAGAGACTAGCTCAGCAGGACACAAACCCGCTATACGC 84261

105 LeuAspAsnIlePheLeuThrArgMetHisTrpSerAsnValGlyGlyLe 121
   |||:::|||||:::|||||:::|||||:::|||||:::|||||
84260 CTGGAACAAATTTTCCTTACACAAATAACATCGGCTTCCTCGCGAGGGTT 84211

121 uSerGlyMetIleLeuThrLeuLysGluThrGlyLeuProLysCysValL 138
   | |||::: |||:::|||||:::|||||:::|||||:::
84210 ACCAGTTTGACGCTCACCATCCAGATCCGCGAGTCCGCGACATTGGAC 84161

138 euSerGlyProGlnLeuGluLysTyrLeuGluAlaIleLysIlePhe 154
   || r |||:::|||||:::|||||:::|||||:::|||||
84160 TCATGAGACCGCTCACTCGGCTCTATGCTGAGTCAATGCGAGCTTC 84111

155 SerGlyProLeuLysGlyIleGluLeuAlaValArgProHisSerAlaPr 171
```


100403 GCGGCTGTGTAGAGGCTGTGGGTGACGTCGGGCTGGCAACTCCTGTCAA 100452
553 553
100453 GAGAGAGGCTGCAGACCCCTAACCCGGAGGGATGSCCCTGGGGCCTGGC 100502
554 Gly. LeuProSerIleLeuLeuG 561
100503 TGACGCATGTCCTCGTGTTCCTTGCAGAGGCTTGCCAAAGTATCTGTGCTGC 100552
561 InArgGluArgAlaLeu..... 566
100553 AGAGAGAACCGCCTTGGTAAGTGTGGCACCTTGATGGCGCTTCTGAGTTT 100602
566 566
100603 CAGCGGTTTACACATCATCCGCCATGCCCTCTTGGCACTCCAGTTTATT 100652
567 Ala. Ser 568
100653 GAGATGTTCTGTCTGAGTCGACACTTGCATTTTGTTCAGGCATCT 100702
569 LeuGlyLysProLeuHisProLeuValAlaProAsnGlnLeuLy 585
100703 TTGGAAAGCCGCTTCACCCCTTTCCTGGTGGTGGTGGCCCAACCAAGCTCAA 100752
585 sAlaTrpLeuGlnGlnTrpHisAsnGlnCysGlnGluValLeuHisHis. 601
100753 AGCCTGGCTCCAGCTACCAACCAACAGTCCAGAGGCTCTGCACCACA 100802
601 601
100803 TCAGGTGAGCATCCAGGGCAGCCTGGCCGCTGGCTGTGTTGCTTGCTGCC 100852
601 601
100853 GTCTCCTTCAGAGCTCAAGGTGGACACTGGGAGTAGTTACCAATATCCCC 100902
601 601
100903 CAGCAGCCTTGCCCTTGACATGGTCCAGATGGCAGNAGCAGGGGAGAAG 100952
601 601
100953 TGCATTGGCTGAAGGACAGAAACCATTAGATAGTTCCCATGTAATGCTTA 101002
601 601
101003 TTTTCTTAGAAGCATTTCTCCAGTCCCTCATTTGAGTCTGAGCTGCTT 101052
601 601
101053 TCTAACTTCGACAGCTTTTCTTGATGACAGACTTCCAGAGCCAAAGCAC 101102
601 601
101103 CCAAAATAGTGGCTAGCACAGAGAATGTCCATAGCAGGTGTGGCTAGCT 101152
601 601
101153 GGCAGGTGGACCATCCTCACCCCAAGGGAAGAGTCCCTCTGTGTTGA 101202
601 601
101203 GCCATCGTGGCGGCTGCTGCCTGAGCGGAGGAGCATTCACCTGTGTTG 101252
601 601
101253 GTTTCCTCCAGTGGCCTAGAGGCTTTGGTTGGCTCTTTATATTGACTG 101302
601 601

101303 CTGTTTCCTCATCATAGTACTATGATTAACTCATGTTTCTCCTAAGA 101352
601 601
101353 ATGATTTTGGGTTCTCCAGCCAAAGACTTAAACTTTGGTTCCAGATGTC 101402
602 Ile. SerMetI1 605
101403 CAAGAAACGTTTATTATCATTTTAAATGTTTGTCTTTTACAGTATGAT 101452
605 eProAlaLysCysLeuGlnGluAlaGluIleSerSerProAlaValG 622
101453 TCCTGCCAAATGCCCTTCAGGAAGGGCTGAGATCTCCAGTCTCGAGTGG 101502
622 luArgLeuIleSerSerLeuLeuArgThrCysAspLeuGluGlu..... 636
101503 AAAGATTGATCAGTTCGCTGTTGGCAACATGTGATTTGGAAGAGGTAAGG 101552
636 636
101553 GGCACAGCCGAGGCATCATGGGGCGAGGTGGGAGCAGAGCTGCAGAG 101602
637 PheGlnThrCysLeuValArgHisCysL 646
101603 CCTCCAGCCCCACCTTTCAGTTTCAGACCTGTCTGGTGGCGCACGCA 101652
646 YshisAlaPheGlyCysAlaLeuValHisThrSerGlyTrpLysValVal 662
101653 AGCATGGTTTGGCTGTGCCCTGTGTGCACACCTCTGGCTGGAAAGTGGTC 101702
663 TyrSerGlyAspThrMetProCysGluAlaLeuValArg..... 675
101703 TATTCGGGGACACCATGCTCGAGGCTCTGCTCCGGATGGGTGAGTA 101752
675 675
101753 GAGGAAGAACAGCCACCCTGAGTTGCTCTGGGTTTGTGTAGCTGGA 101802
675 675
101803 GGTCAATGCAGTGGGCTTGCAGGGAACGTCACGAGGAGGAGACTC 101852
675 675
101853 AGTCCCCACCTCAGAGTCTCTGGTTGTCATCCTAGTAGGACAGCCAG 101902
675 675
101903 GGCAGGGGAGCTGAGTTTGAGACCAGGAACACGACGACTGAGGCC 101952
676 Met 676
101953 TGCTGCGGCTCTCGCAGAGAACTCTGCCCTGATCCTTGTGCTGCTCTC 102002
677 ..GlyLysAspAlaThrLeuLeuIleHisGluAlaThrLeuGluAspGly 692
102003 CAGGAAAGATGCCACCTCTCTGATACATGAAGCACCCCTGGAAGATGGT 102052
693 LeuGluGluGluAlaValGluLysThrHis..... 702
102053 TTGGAAGAGGAAGCAGTGGAAAGACACACAGAGTAGCAAGGCCGGTCTAG 102102
702 702
102103 TCCTTTCGCCCCACATCCTCTCCCTCCCCACTACGTGACACTGAGCAGC 102152
702 702
102153 CGTGGTTTGTCTCCACTGATGTGGGGCTGCCCTGCTTCTTATCAAGGCT 102202
702 702
102203 ATGGGGGCTTCTTGACCTGTGGCAGTGCTCACAGGCTCTTGCGCTTTAT 102252

```
702 ..... 702
102253 TTTTCAGAAATTTCTAAGCAAGATTCTAGAGTGAGGCACAGTTTTTTGA 102302
702 ..... 702
102303 AAGCATCTAGAAATCGGCTGAATAAACAATAAAGCCATGTACAGGAATTCG 102352
702 ..... 702
102353 CAGGGGAAGCGGGGCTGGGGACTGAATTTTGGCTGCTAATTTCAAC 102402
702 ..... 702
102403 GAAAGAGTGCATTACCCAGGTGGGCCCTGTGTTCTCTGGGTGCCCT 102452
702 ..... 702
102453 CATGGACAGATTTGGAGCCACACAGAGGGTGGGCTTCATCCAGGGTG 102502
702 ..... 702
102503 TGTGCCAAGGCTTGGCCCTCAGGGGAGATTGTGTGGCTACGGAGGTGC 102552
702 ..... 702
102553 CGGTTAAGAAAACCCACCAGCTTCCCGGGTGCCCTGGCAGTTGATGGCC 102602
702 ..... 702
102603 AGGGTCTGGCCACTGTCTGCTTTCAGTCTTGCAGTCTTGCAGTTCAGCTTC 102652
703 ..... Ser...ThrThrSerGlnAlaIleSerVal 711
102653 AGTCTGCTCTGCTTTCACCTGCACACACACCTCCCAAGCCATCAGCGTG 102702
712 GlyMetArgMetAsnAlaGluPheIleMetLeuAsnHisPheSerGlnAr 728
102703 GGGATGCGGATGAAGCGGAGTTCATTATGCTGAACCACTTCAGGCAGCG 102752
728 gTyrAlaLysValProLeuPheSerProAsnPheSerGluLysValGlyV 745
102753 CATGCCAAGGTCCCTCTTCAGCCCCCACTTCAGCGAAGAGTGGGAG 102802
745 aIAlaPheAspHisMetIys...Val..... 752
102803 TTGCTTTTGACCATGAAGGTCTGTATGTCACACGACAGCACAGGGCG 102852
752 ..... 752
102853 GGGACGGGCGAGGAGACAGGACTTACACACTGAGTAGGACGGTCAGCT 102902
752 ..... 752
102903 GGAGTTTGTCTTATTATTGGGGCCACCGTGGGAAAAGTTATACCCA 102952
752 ..... 752
102953 TCACCTAACCAAGTTCGAACACACCTGGGTTTGTGTTGAGACCCACCTCCT 103002
752 ..... 752
103003 GCAGGGGCCAACTAGTCTTCACTCTCAGTTCACCTGGAAATTTCTGAGAAT 103052
752 ..... 752
103053 CCTTTTAGCCCTGGACTGCTCACACAGTCATGGCAATTCAGCCCTCAGCAC 103102
752 ..... 752
103103 AGACCTGTGACAGAGGTGGTTGCCCTTGTGTGAGTGGGAAAGCCAGGCCTG 103152
```

```
752 ..... 752
103153 ACCCTTGGCCTTCCGGAATGAAGGGCAGAGCGGAGCCAGGCGCTCGTTT 103202
752 ..... 752
103203 TTCAGAGCTTGTATTTTGAGAGCATCTGGACTGCTCTCCCTTCCCTCTCC 103252
752 ..... 752
103253 GGAGGCCCTTAGCCAGGCTGGGAGCCCTTGCCTTTAGAGGGTTCCC 103302
752 ..... 752
103303 TCCATGCCATTCTTTTTCATTTCAGCTGTGGCCTGTGGTGTGGCCA 103352
752 ..... 752
103353 AGGAAGGGCGTTGGCGTGTGTGTGACACATGACTGCATCCCTTCCA 103402
752 ..... 752
103403 GCTCTGTCCCCCACCCCTGCCCCCTGTGACACATGCTCTTGTCTTATT 103452
753 ..... CysPheGlyAspPheProThrMetProLysLeuIle 764
103453 GTGTCTTCTAGGTCTGCTTTGGAGACTTTTCCAACAATGCCCAAGCTGATT 103502
765 ProProLeuLysAlaLeuPheAlaGlyAspIleGluMetGluGluAr 781
103503 CCCCACACTGAAAGCCCTGTTTCTGTCGCACATCGAGGAGATGGAGGCG 103552
781 gArgGluLysArgGluLeuArgGlnValArgAlaAlaLeuLeuSerArg 798
103553 CAGGAGAAGCGGAGCTGCGCAGGTGGGGGGCCCTCTCTGTCCAGGG 103602
798 luLeuAlaGlyGlyLeuGluAspGlyGluProGlnGlnLysArgAlaHis 814
103603 AGCTGGCAGCGGCTTGGAGATGGGAGCTCAGCAGAAGCGGGGCCAC 103652
815 ThrGluGluProGlnAlaLysLysValArgAlaGln 826
103653 ACAGAGGAGCCACAGGCAAGAGGTACAGGCCCG 103688
seq_name: gb_pl3:SPAC1D4
seq_documentation_block:
LOCUS SPAC1D4 31201 bp DNA PLN 15-OCT-1999
DEFINITION S.pombe chromosome I cosmid clD4.
ACCESSION Z69239
VERSION 269239.1 GI:1177333
KEYWORDS byr1; cct2; CDP-alcohol phosphatidyltransferase; chaperonin;
conjugation and sporulation; csk1; cyclin suppressing protein
kinase; DNA repair helicase; golgi peripheral membrane protein;
protein kinase; rad15; rhp3; serine threonine protein kinase; stel;
tcp1 beta subunit homologue; tf2 LTR.
SOURCE
ORGANISM Schizosaccharomyces pombe
Eukaryota; Fungi; Ascomycota; Schizosaccharomycetales;
Schizosaccharomycetaceae; Schizosaccharomycetes.
REFERENCE
AUTHORS 1 (bases 1 to 31201)
TITLE Direct Submission
JOURNAL Submitted (01-FEB-1995) Schizosaccharomyces pombe chromosome I
sequencing project, Sanger Centre, Hinxton Hall, Hinxton, Cambridge
CB10 1RQ E-mail: barrell@sanger.ac.uk
Notes:
Details of yeast sequencing at the Sanger Centre are available on
the World Wide Web.
(URL, http://www.sanger.ac.uk/Projects/S\_pombe)
Protein coding regions (CDS) have been predicted with the help of
computer analysis using the Genefinder program in PomBase (an ACEDB
database) with additional predictions for the branch-acceptor sites
```

supplied by the program Sp3splice. CAUTION: It is possible that for any individual CDS we may have underestimated or overestimated the number of introns/exons or we may not have chosen the correct splice donor/acceptor sites.

CDS are numbered using the following system eg SPAC5H10.01c. SP (S. pombe). A (chromosome 1), c5H10 (cosmid name), .01 (first CDS), C (complementary strand).

The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous.

The length in codons is given for each CDS.

IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions.

Cosmid c1d4 is overlapped at the 5' end by cosmid c1f3 and at the 3' end by cosmid c22f3.

FEATURES

source

```
1. 31201
   Location/Qualifiers
     organism="Schizosaccharomyces pombe"
     strain="972h-"
     db_xref="taxon:4896"
     chromosome="I"
     clone="cosmid c1d4"
     map="IL"
1. 133
   misc_feature
```

```

   /note="the sequence of this region is derived from cosmid
   c 1f3, the true end of c1d4 is at position 134, while the
   true end of c1f3 is at position 4137 in this sequence"
7. 864
```

```

   /gene="SPAC1f3.11"
   /note="SPAC1D4.01"
7. 864
```

```

   /gene="SPAC1f3.11"
   /note="SPAC1D4.01, len:285"
   /codon_start=1
```

```

   /product="hypothetical 33.0 kd protein"
   /protein_id="CAB56125.1"
   /db_xref="GI:5912524"
```

```

   /translation="MRQIFGCTGMYGIFQEGLEIPPTTKSIRMNSIHKKKSNRS
   FRRKVFGELELDNDIRLQALEATKRRKIRNSIIGINAELKLLNQETKKEK
   QLTANPEHNEADOTSQSSKLI EAOLPTVEDRAKOTNEVDINTHLLNFVEKIKQE
   RL AONYSENGENTNALNKNESTVQNIKNSLHPNESHSTRDAAALGATREVDLGIISTD
   VDLNKRGRKKRKARKEKLDLSRALRTSEDAARDEFIEKRLKPLSQDEESKGIYRRF
   RVYKDGTDQ"
   complement(join(1088..1636,1686..1873,1970..2090,
   2137..2316))
   /gene="SPAC1D4.02c"
   complement(join(1088..1636,1686..1873,1970..2090,
   2137..2316))
   /gene="SPAC1D4.02c"
   /note="SPAC1D4.02c, len:345, SIMILARITY:Rattus norvegicus,
   O35254, golgi peripheral membrane protein p65., (451 aa),
   fasta scores: Opt: 493, E():4.7e-20, (34.4% identity in
   305 aa)"
   /codon_start=1
   /label="SPAC1D4.02c"
   /product="putative golgi peripheral membrane protein"
   /protein_id="CAB56126.1"
   /db_xref="GI:5912525"
```

```

   /translation="MFGGLKNFIEKSEALAGIHRESDSCGFRVLKVENDSKAYNAR
   IESTYDFITANGILLGDPFMFALLRSDSPVETLVFSIKGOITRKVNKINSDEK
   IGMVLQWASTAPADALHILNVDDSPARASLVPYEDYIVGTPEGMTGKALSDL
   IESHLNPLRLIYINHYRDSRTQVTIVPNRHGNGAIGCGVGHVLRHLPAPLSGPP
   POPGDIVFSPMLGDPHKVSPSETENFLTPRPPKIASANAGSSNIEISIPHYORHK
   KSHKATODSSIQSYLDEEKLRELDHKTKDASNDTSQTPLPPPPVAVNSTNDE
   SAPONEELVKN"
   complement(1637..1651)
   /note="splice branch and acceptor sequence,
   ctaacaattacaag"
   complement(1680..1685)
   /note="splice donor sequence, gtagtgt"
   complement(1874..1886)
   /note="splice branch and acceptor sequence, ctaacgactacag"
```

misc_feature

```

   /note="splice branch and acceptor sequence,
   ctaacaattacaag"
   complement(1680..1685)
   /note="splice donor sequence, gtagtgt"
   complement(1874..1886)
   /note="splice branch and acceptor sequence, ctaacgactacag"
```

misc_feature

```

   /note="splice donor sequence, gtagtgt"
   complement(1874..1886)
   /note="splice branch and acceptor sequence, ctaacgactacag"
```

misc_feature

```

   /note="splice branch and acceptor sequence, ctaacgactacag"
```

misc_feature

```

   /note="splice donor sequence, gtagtgt"
   complement(2091..2108)
   /note="ctaaagtgtttcacag, splice branch and acceptor"
```

misc_feature

```

   complement(2131..2136)
   /note="gtagtg, splice donor sequence"
```

gene

```

   /gene="SPAC1D4.03c"
   complement(join(2574..2820,2907..4201))
```

CDS

```

   /gene="SPAC1D4.03c"
   complement(join(2574..2820,2907..4201))
```

```

   /note="SPAC1D4.03c, len:513, SIMILARITY:Fugu rubripes,
   O9YGN1, sand protein., (520 aa), fasta scores: opt: 609,
   E():7.2e-32, (28.5% identity in 407 aa)"
   /codon_start=1
   /label="SPAC1D4.03c"
   /product="hypothetical protein"
```

```

   /protein_id="CAA93212.1"
```

```

   /db_xref="GI:1177336"
```

```

   /db_xref="SWISS-PROT:Q10150"
```

```

   /translation="MEPTSEHSIKKEEVNVRHSESGCLLLNPGNVLMAAPSV
   SEDQEVSRSTPELRSHVENVEQLLSLDHNSPLNTSVSSSSNTAVDEIKLL
   SLISFDLAKQRKTYLIFSSGKPVFSNIVDDSIETPTVGALQAIISSEVSKKEETSF
   STFSNVIVLVSKNPLYLVGSPSTTLSAAYLLSELNLLYCOLLTGVTAKMOLTNRSR
   PNFDLRLIGSNEQFLKELQDLNDYELVPLTNAISPLRASSFPDQSQLLRTPPK
   SLLEFTTAIRGLVCMVRKAKLLHANDLYLLFSLFRQSFNDSMEHWVPVCFPTLN
   PDAYIYISYFLCKDVTILIMSGSEGVFQSVKCAQVQIQDHWKLLKLYCEMD
   RTTPRNPSCISHYLFYSKYSQFTPGYSFSTPNFNTRLTYAIYASLHDOAFHKKN
   SFSINMTVHESLLLTWTSTADFHCIANATSSOLLIANVKNILRWIRRENRLFQI
   TNLSP"
```

```

   complement(2821..2836)
   /note="splice branch and acceptor sequence,
   ctaactgttttta g"
```

```

   complement(2901..2906)
   /note="splice donor sequence, gtagtgt"
```

gene

```

   join(5400..5402,5439..5570,5618..7066)
   /gene="SPAC1D4.04"
```

CDS

```

   join(5400..5402,5439..5570,5618..7066)
   /gene="SPAC1D4.04, len:527"
```

```

   /codon_start=1
```

```

   /product="probable t-complex protein 1, beta subunit"
```

```

   /protein_id="CAA93213.1"
```

```

   /db_xref="GI:1177337"
```

```

   /db_xref="SWISS-PROT:Q10147"
```

```

   /translation="MSLNPHOIFNESGIERQENARLSFVGAIAGDLVKSTLGPKG
   MDKLIQSNSSGDIYVNDGATILKSIALDNRAAKVLVNIKVDQDEVGDGTSVCVFA
   AELLQAEIMVNAKIHPOVIIDGRIATKTAIDLRASSIDNSSDPKFRSDELIAR
   TIGVCPKVMENANILIAVMDTKVFGARVVDVTGKLAELERAEKRMKAKVE
   KIKSHINCFINRQLIYNPQOLFADAGTMSIEHADFDGIERLSLVTGTEIATFDHP
   ELVKLGHCKKTEIITIGEDKMTKFSGVGAEACTIVLRGATHOLLDESERATHDALAV
   LSQVVAESRVTILGGCAEMLKAKAVEEATPEPGKAVAVSAFALKSQLPILADNA
   GFDSSELVAQLKAAHYDNGDNTGLMDDEGATDMRAKGLILEALKLQKQAVSSGSEGAQ
   LLRLRVDTILKAAAPRRER"
```

```

   5425..5438
   /note="splice branch and acceptor sequence,
   tactaacacatatag"
```

```

   join(5529..5570,5618..7045)
   /gene="SPAC1D4.04"
```

misc_feature

```

   /note="Match to PF00118 cpn60_TCP1, TCP-1/cpn60 chaperonin
   family Score 590.77"
```

misc_feature

```

   5571..5576
   /note="splice donor sequence, gtagtgt"
```

misc_feature

```

   5604..5617
   /note="splice branch and acceptor sequence,
   ttaacgatttatag"
```

misc_feature

```

   5741..5767
   /gene="SPAC1D4.04"
```

gene

```

   /note="PS00995 Chaperonins TCP-1 signature 3"
```

CDS

```

   complement(join(7282..8366,8419..8497))
   /gene="SPAC1D4.05c"
   complement(join(7282..8366,8419..8497))
```

```

/gene="SPAC1D4.05c"
/notes="SPAC1D4.05c, len:387, LOW SIMILARITY:Mus musculus,
Q920U0, polytopic murine leukemia virus receptor sygl.,
(695 aa), fasta scores: Opt: 261, E():1.5e-10, (25.1%
identity in 398 aa) also shows low similarity to,
Saccharomyces cerevisiae, ERD1_YEAST, ERD1 protein., (362
aa), fasta scores: opt: 187, E(): 1e-05, (23.6% identity
in 382 aa)
SPAC1D4.05c, len:387, SIMILARITY:Mus musculus, Q920U0,
polytopic murine leukemia virus receptor sygl., (695 aa),
fasta scores: Opt: 261, E():1.5e-10, (25.1% identity in
398 aa)"
/codon_start=1
/label=SPAC1D4.05c
/product="hypothetical protein"
/protein_id="CAA93214.1"
/db_xref="GI:1177338"
/db_xref="SWISS-PROT:Q10151"
/translation="MDLEVVEFPLHKKLALPFRIGLLVIVGTWLSVCYHLIYLNR
QPISPRGSLSNKKWHLQIPLSNRHTDLEENTERKANLIVSPVDFHAGYCFALILSI
SWATGFILEKKTGGDIGGYSHPYIPLWVITAFILVFPFWRYSQRGLKSGII
RVFLQADPRSPYKDFIVSEIFTYAKALGDYIFGVLQGHISKFTLPDLKCDGT
FVPLAMAYPFIIVAILQCLHYGLSRKHTFKINLLSALKHATALPVYLSAIIHAKOT
KFTLTSGHGYLFWLWLSALLSSAYTFIWDVFIWMRIRPFPHKSINHKRPMEIYALG
CFINFLIRVTSWKLHPRLHQPHEYEMGFSEFMELIRFLWLFHDLDAISS"
/misc_feature
complement(8367..8389)
/notes="splice branch and acceptor sequence,
tactaacaactctcta tcttttag"

```

```

alignment_scores:
  Quality: 785.50      Length: 852
  Ratio: 1.650        Gaps: 32
  Percent similarity: 55.869  Percent identity: 28.404

alignment_block:
US-09-434-382-2 x SPAC1D4

Align seg 1/1 to: SPAC1D4 from: 1 to: 31201

59 TyrLeuGlnValAlaAlaGlySerArgAspSerGlyAla.....Al 73
|||||  |||:|||||:|||||:|||||:|||||:|||||:|||||:
14502 TATTACAAATTTGTGAGCGTCAGTCTCGGATACATCTGTATTCCTCG 14551

73 aleuTyrValpheSerGluPheAsnArgTyrIleuPhe...AsnCysGlyG 89
|||||  |||:|||||:|||||:|||||:|||||:|||||:
14552 CATCCACTTATTTTGTGATCTTAAGAGGTATGTTTGAAGCGTAGGGG 14601

89 luGlyValGlnArgLeuMetGlnGluHisLysLeuLysValAlaArgLeu 105
|||||  |||:|||||:|||||:|||||:|||||:|||||:
14602 AAGGTTGTCAAGAGCTATCTCTCCACAGCTTCGATTATCAAGATT 14651

106 AspAsnIlePheLeuThrArg..... 112
|||||  |||:|||||:|||||:|||||:|||||:
14652 AAAGATGCTCTTCTTAATCAAGAGCTTCATATCATCACCAGATACATA 14701

112 ..... 112
14702 TGATTTCTTCTTCTCATCGTCTACTACCTCTGTTTCTGATGCTACAGC 14751

112 ..... 112
14752 TGGATGATCGGACAAAGTTATTATGTATCTGAACGTAATAGTATGTGCTCG 14801

113 .....MethisTirpSerAsnValGlyGlyLeuSerGly.. 123
|||||  |||:|||||:|||||:|||||:|||||:
14802 ACTGTAAACTACTCTACATGCTGGGATCTTGTGCTGGGTTCCAGGTAA 14851

124 .....MetIle.LeuThrLeu..... 128
|||||  |||:|||||:|||||:|||||:|||||:
14852 GTGTTAAATATATTAAATTAATTTCTGACAAATAGGCTTTTGTGCTCTC 14901

129 .....LysGluThrGly...LeuProLysCysValIle 138
|||||  |||:|||||:|||||:|||||:|||||:

```

```

14902 AATGATATTTCGGAACCTCGGGAAACCGGAGAGCTAGTCCTTTTGTCT 14951
138 uSerGlyProGlnLeuGluLysTyrLeuGluAlaIleLysIlePheS 155
|||||  |||:|||||:|||||:|||||:|||||:|||||:
14952 TCATGACCTAGTGAAGTTCATCAGCTTTTATCTTCGATCGCGCATTTCA 15001
155 erGlyProLeuLysGlyIleGluLeuAlaValArgProHisSer..... 169
|||||  |||:|||||:|||||:|||||:|||||:
15002 CTTATCAT...ACGAATGTAATTTAACGGTACAGGATATATCTTCGGCT 15048
170 ...AlaProGluTyrGluAspGluThrMetThrValTyrGlnIleProI 185
|||||  |||:|||||:|||||:|||||:|||||:
15049 GAGCCCCCAGTATTCGTTGATGAAATATTTGTAACTCCGCTAGTTGT 15098
185 e.....HisSerGluGlnArgA 191
|||||  |||:|||||:|||||:|||||:
15099 TTCACCTAGTTAAATTTCTTTTAAAAAAGGAGCATGATAAATCAACA 15148
191 rgGlyLysHisGlnProTrpGlnSerProGluArgProLeuSerArgLeu 207
|||||  |||:|||||:|||||:|||||:|||||:
15149 GAGGTACAAAT.....GCAAGACCATTAAGAG... 15177
208 SerProGluArgSerSerAspSerGluSerAsnGluAsnGluProHisLe 224
|||||  |||:|||||:|||||:|||||:
15178 .....GATAGAGCAATACTTCT.....CCTCATTTG 15203
224 uProHisGlyValSerGlnArgArgGlyValArgAspSerSerLeuVal 241
|||||  |||:|||||:|||||:|||||:|||||:
15204 GTATTCTCACGTTCC.....ATGATACATCATCTTCGTCG 15238
241 al..... 241
|||||  |||:|||||:|||||:|||||:
15239 TTGAGATGCAATGTATAACACTCGCGCTCTTTTGGAAACGACAGCA 15288
242 .....AlaPheIleCysLysLeuHisLeuLysArgGlyAsnPh 254
|||||  |||:|||||:|||||:|||||:|||||:
15289 GAGCTGTTTATTTCTTATTTATTTCAATCCCATCTACCCAGGAAAT 15338
254 eLeuValLeuLysAlaLysGluMetGlyLeuProValGlyThrAlaAlaI 271
|||||  |||:|||||:|||||:|||||:|||||:
15339 TCACGCCCAAGCAAAATCGCTTGGTATCACGAAAGGATTAGACTGT. 15387
271 leAlaProIleleAlaValLysAspGlyLysSerIleThrHisGlu 287
|||||  |||:|||||:|||||:|||||:|||||:
15388 .....GGTAGACTCGCTCGAGGAGAACCTGTAACCTCTTGA 15423
288 ...GlyArgGluIleLeuAlaGluGluLeuCysThrProAspProGl 303
|||||  |||:|||||:|||||:|||||:|||||:
15424 AACGAAACAGTATATCCAAAAGAGTGATTTGCTCTTCTTATTCCTGG 15473
303 yAlaAlaPheValValValGluCysProAspGluSerPheIleGlnProI 320
|||||  |||:|||||:|||||:|||||:|||||:
15474 CTCCTCTTTTATTTATTTATTTGCTGCCCCAAGCACTCGTAATTCGCTGG 15523
320 leCysGluAsnAlaThrPheGlnArgTyrGlnGlyLysAlaAspAlaPro 336
|||||  |||:|||||:|||||:|||||:|||||:
15524 TCATTGAAATCATAAATGGACA.....AATGCACCTAAACCT 15561
337 ValAlaLeuValValHisMetAlaProAlaSerValLeuValAspSerAr 353
|||||  |||:|||||:|||||:|||||:|||||:
15562 GTTGT...GTCATTTCATTCGTTACACCAAGATATACAAAATCCACG 15608
353 gTyrGlnGlnTrpMetGluArgPheGlyProAspThrGlnHisLeuVal 370
|||||  |||:|||||:|||||:|||||:|||||:
15609 TTATCAGAGCTGGATTTCTTTCATCCCAAGTGAGGTGAGCCATCTAATG 15658
370 euAsnGluAsnCysAlaSerValHisAsnLeu...ArgSerHisLysIle 385
|||||  |||:|||||:|||||:|||||:|||||:
15659 CTTGACTGAAGTAAACGAAGTAAATAATATCTTAGAGTGCAGCTCGCT 15708
386 GlnThrGlnLeuAsnLeuIleHisProAspIlePheProLeuLeuThrSe 402
|||||  |||:|||||:|||||:|||||:|||||:
15709 ATTGCACACACTTAATTTGTTAGATTCAAAAGTTTTTCTCTTTA.....GG 15752

```

```

402 rPheArgCysLysLysGluGlyProThrLeuSerValProMetValGlnG 419
:|||||
15753 CTTTAACTGC.....TATGAAGTTAAATAATGACAAA 15784
:|||||
419 LyGlu.....CysLeuLeuLysTyrGlnLeuArg.....Pro 429
:|||||
15785 AAAACAANTCGAATTCGTTTGGCAACCCAACTCGCTTTCGGTTTGA 15834
:|||||
430 ArgArgGluTrpGlnArgAspAlaIleThrCysAsnProGluGluPh 446
:|||||
15835 AAAAAAAGCTGGTATTGACGATTGCGAAGTTGGGTTTCGATTGAAGCTT 15884
:|||||
446 eIleValGluAlaLeu.....GlnLeuProAsnPheGlnGlnSerValGlnG 462
:|||||
15885 GAAGGATAAAATTTTGAAGAGAAACCAAGATTATAAATCATTTGTTGAAG 15934
:|||||
462 lUtyrArgArgSerAlaGlnAspGlyPro...AlaProAlaGluLysArg 477
:|||||
15935 AGCTCAAAAATATGTTAGTAGATAACCAAGCCAGCTTCAGGT 15984
:|||||
478 SerGlnTyrProGluIleIlePheLeuGlyThrGlySerAlaIleProMe 494
:|||||
15985 TCTGACATC...CAAAATTGTACTCTTGGCACTGGATCGCAATGCCITC 16031
:|||||
494 tIlyIleArgAsnValSerAlaThrLeuValAsnIleSerProAsp.... 509
:|||||
16032 TTTATACAGAAATGTTTCATCTACTTACGTTAGAAATTCGCTTGACAAA 16081
:|||||
510 .....ThrSerLeuLeuLeuAspCys 516
:|||||
16082 AATGTATGGAAGACATCAGCTATCATGTATGAAACATACATTTGGACTGT 16131
:|||||
517 GlyGluGlyThrPheGlyGlnLeuCysArgHisTyrGlyValAspGlnValAs 533
:|||||
16132 CGAGAGGGCACTTTGGGAGCTCTATCGAGGAGTATGGTGAATTTAA 16181
:|||||
533 pArgValLeuGlyThrLeuAlaAlaValPheValSerHisLeuHisAlaAla 550
:|||||
16182 APACGAAATAGCTTTTACGATGGATTATATCTCCCATATGATGCTG 16231
:|||||
550 sPHisHisThrGlyLeuProSerIleLeuLeuGlnArgGluArgAlaLeu 566
:|||||
16232 ACCATCATGCTGCTGTATAGGGTTTTTA.....AAGCTGG 16269
:|||||
567 AlaserLeuGlyLysProLeuHisProLeuLeuValValAlaProAsnG1 583
:|||||
16270 ACGAAGTACTCAGATCGGAGATCAAGTTGTTTATAACAGCACCTCCACA 16319
:|||||
583 nLeuLysAlaTrpLeuGlnGlnTyrHisAsnGlnCysGlnGluValLeuH 600
:|||||
16320 ATTGAATTTTGGCTACTAGATACTCAAGGATTGATTATCTACCATTTAT 16369
:|||||
600 iSHisIleSerMetIleProAlaLysCysLeuGlnGluAlaGluIle 616
:|||||
16370 CCAACATCGTATTTATTTCCAAATAGTCTTTACGTACTGACAGGAGCGC 16419
:|||||
617 SerSerProAlaValGluArgLeuIleSerSerLeuLeuArgThrCysAs 633
:|||||
16420 TCTGCTTGGATCCTCTAGGTTG...TCTTCTTTATTTAAAGAAATTTGA 16466
:|||||
633 pLeuGluGluPheGlnThrCysLeuValArgHisCysLysHisAlaPheG 650
:|||||
16467 TTTAGTATCTTTTGAACGCTTCTGCCATACACTGCCGCTATTCATATT 16516
:|||||
650 lYcysAlaLeuValHisThrSerGlyTrpLysValValTyrSerGlyAsp 666
:|||||
16517 GCATGGAGATTACTAATTCACGTGGTTGGAAGATCGCATACTCTGGAGAT 16566
:|||||
667 ThrMetProCysGluAlaLeuValArgMetGlyLysAspAlaThrLeuLe 683
:|||||
16567 ACTCGACCTTCTGAGGATTTTCTGATACATTTGCCAAGACCTCACTTTATT 16616
:|||||

```

```

683 uIleHisGluAlaThrLeuGluAspGlyLeuGluGluAlaValGluL 700
:|||||
16617 AATTCAGGAAGTACTTTGGAGACTCTATGCAGAAATTCGAATAAAA 16666
:|||||
700 ySThrHisSerThrThrSerGlnAlaIleSerValGlyMetArgMetAsn 716
:|||||
16667 AGCAACACAGTACTTACTCTGAAGCTTTGAAGTGGCCAAAAAAGCTGGT 16716
:|||||
717 AlaGluPheIleMetLeuAsnHisPheSerGlnArgTyrAlaLysValPr 733
:|||||
16717 ACAAAAAACGTTATTATTGACTCATTTTGGCAACGATATCCCAAAATACC 16766
:|||||
733 oLeuPheSerProAsnPheSerGlu...LysValGlyValAlaPheAsph 749
:|||||
16767 AGACATTGATATTTTCGACAGAGACTTACATATAGCTTTGGCGTTTCATG 16816
:|||||
749 iSMetLysValCysPheGlyAspPheProThrMetProLysLeuIlePro 765
:|||||
16817 GAATGACTCTTAAATATCGGATATCTCATTTGTTCTGTTATTTTGGAAAG 16866
:|||||
766 ProLeuLysAlaLeuPheAlaGlyAspIleGluGluMetGluGluArgAr 782
:|||||
16867 CCTCTGGCTTATCTATT.....AATGAGAAATTTAAAGAGGAGTC 16910
:|||||
782 gGlu 783
:|||||
16911 GGAT 16914
:|||||
seq_name: gb_pl1:AC022354
seq_documentation_block:
LOCUS AC022354 46335 bp DNA PLN 01-FEB-2000
DEFINITION Arabidopsis thaliana chromosome I BAC F915 genomic sequence,
complete sequence.
ACCESSION AC022354
VERSION AC022354.1 GI:6850338
KEYWORDS HTG.
SOURCE
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
Brassicales; Brassicaceae; Arabidopsis.
Lin,X., Kaul,S., Town,C.D., Benito,M., Creasy,T.H., Haas,B., Wu,D.,
Bowman,C.L., White,O., Fujii,C.Y., Utterback,T.R., Barnstead,M.E.,
Ronning,C.M., Koo,H., Nierman,W.C. and Fraser,C.M.
Arabidopsis thaliana chromosome I BAC F915 genomic sequence
Unpublished
2 (bases 1 to 46335)
Direct Submission
Submitted (01-FEB-2000) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA
Address all correspondence to:atetigr.org

```

BAC clone F915 is from Arabidopsis chromosome I and is near the molecular marker m213. The orientation of the sequence is from SP6 to T7 end of the BAC clone. Genes were identified by a combination of three methods: Gene prediction programs including GRAIL (available by anonymous ftp from arthur.epm.ornl.gov), Genefinder (Phil Green, University of Washington), Genscan (Chris Burge, <http://gnomic.stanford.edu/~chris/GENSCANW.html>), and NetPlantGene (<http://www.cbs.dtu.dk/netgene/cbsnetgene.html>), searches of the complete sequence against a peptide database and the Arabidopsis EST database at TIGR (<http://www.tigr.org/tdb/at.html>). Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as 'unknown' proteins. Genes without protein or EST similarity, that are predicted by more than two gene prediction programs over most of their length are

annotated as 'hypothetical' proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats are identified by repeatmasker (Arian Smit, <http://ftp.genome.washington.edu/RM/RepeatMasker.html>).

FEATURES

Location/Qualifiers

```

source
1..46335
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="1"
/map="m213"
/clone="F915"

misc_feature
1..5778
/notes="overlap with BAC clone F5F19
(AC006216:105116..110893)."
```

```

1
/notes="53772 nt beyond this point were not included in the
submitted sequence due to an overlap with another BAC
(F5F19)"
complement(1497..1591)
/rpt_family="(TAAAA)n"
complement(2125..2194)
/rpt_family="(GAAA)n"
complement(join(<7284..7610,7681..8652,8869..9016,
9063..9373,9467..9826,9904..9947,10041..10095,
10177..10221,10306..10395,10478..>10639))
/genes="F915.1"
complement(<7284..>10639)
/genes="F915.1"
/notes="predicted by genefinder"
complement(join(7284..7610,7681..8652,8869..9016,
9063..9373,9467..9826,9904..9947,10041..10095,
10177..10221,10306..10395,10478..>10639))
/genes="F915.1"
/notes="hypothetical protein"
/codon_start=1
/protein_id="AAF29402.1"
/db_xref="GI:6850339"

translation="MENNEATNGSKSSNSFVFNKRRAFGPDITOKKRNKLSDKL
NPTNTIAYAQILGTGTDQTSSTSVLLFFDKQRFIFNAGEGLQRCTEHKIKLSKD
VFLSVCSSETAGLPGLLTLTAGIGEGLSVNVGPSDNLVLDAMKSEIPRAAMVHT
RSFGSSPDPIVNVDEVKLSAIIKPCHEEDSGNKGSLSVYVVCELPEILGKGF
DLEKAKVPGKPSYSLQSGSESDERDITVHPSDVMGSPISPGIIVLLVDQPTF
SHAELFLSKLESYSPDEQITGAKFVNCIIHLSPPSSPTSPYQSWMKFHLTQH
ILAGHORFLPLIIIVSHQKTVRKNNAPILKASRIAARLNLCPQFPAPQFWPSOL
TNSIIDTPPSKNELRPVIRGIDRSCIPAPLTSSVVVDELLSEIPEIKDKSEELKQ
FNKQHNKTIIEKLNLSECVLPNCLEKIRDDMEIVILGTGSSQPSKYRNVSAIFI
DLFSGSLLDGEGTLGOLKRRYGLDGADEAVRKLRCIWIISHIHADHTGLARIAL
RSKLKGVTHPEVIVVGPRLKRFIDAYQRLEDLDMFELDCRSTTATSWASLESGEA
EGSLTQSPMSQSVFKRSDISMDNSVLLCLNKKVLSEIGLNDLISFPVYHCQAY
GVVTKAARVNSVGQIILGWKMVYSGDSRCPETVEASRDATILITHEATFEDALIEA
LAKNHSTTKEADVGSAANVYRIVLTHFSORYPKIPVIDESHMHNTCTAFDLMSINA
DLHLVPKLPYEKTLFRDEMVEDEDADDVAMDDLKEAL"
10641..10753
/rpt_family="(GAA)n"
10928..11001
/genes="F915.2"
/product="trna-Ala"
/anticodon="(pos:10962..10964,aa:Ala)"
10928..11001
/genes="F915.2"

repeat_region
complement(27450..27491)
/protein_id="AAF29405.1"
/db_xref="GI:6850342"

translation="WMRYCQFTPYDPSSNMDDLKAFEDETWKYKKKEAWSLFDEVLY
SEEMKNMGYTLNIMVNECFKIGRDOAIQIFVKTALKKPNPDVTCYRNITTLNKO
GVLFPAEYFFEEMCSDRLYPDVFTYTTMIDAYLKAGKTEDALRISNKKWDAFLQVA
WLACL"
complement(27450..27491)
/rpt_family="(TA)n"
complement(join(<29019..29311,29402..29473,29582..>29780))
/genes="F915.6"
complement(<29019..>29780)
/genes="F915.6"
/notes="similar to delta tonoplast integral protein
GB:AAC49281 from [Arabidopsis thaliana]"
complement(join(11317..11583,12194..>12301))
/genes="F915.3"
/notes="F915.3"
/notes="similar to delta tonoplast integral protein"
GB:AAC49281 from [Arabidopsis thaliana]"
complement(join(11317..11583,12194..>12301))
/genes="F915.3"
/product="putative delta tonoplast integral protein"
/codon_start=1
/protein_id="AAF29403.1"

/db_xref="GI:6850340"
translation="NAVDEVSIIFIGVASLTVSWWFINKFTYLEAVPTSNAIPIHSA
VRVGSQTVRMELIITFALVTVYVATAIDSNNGTCTIAPLAILRVGANILAAGPFS
GGMPPGRSFGSSLAGVNFSGH"
13176..13409
/notes="novel repeat from T25M19 22823 to 23610 788 nt"
13177..13298
/notes="novel repeat from T25M19 22823 to 23610 788 nt"
15264..15339
/rpt_family="(GAAA)n"
complement(18843..18918)
/rpt_family="(GAAA)n"
19668..19760
/rpt_family="POLY_A"
complement(20890..20985)
/rpt_family="(GA)n"
complement(20931..20985)
/rpt_family="(GAA)n"
join(<21044..21140,21233..21450,23180..23739,24098..25199)
/genes="F915.4"
<21044..25199
/genes="F915.4"
/notes="similar to PEPTIDE TRANSPORTER PTR2-B GB:P46032
from [Arabidopsis thaliana]"
join(21044..21140,21233..21450,23180..23739,24098..25046)
/genes="F915.4"
/codon_start=1
/product="putative peptide transporter"
/protein_id="AAF29404.1"
/db_xref="GI:6850341"

translation="MENPPNETEAKQIQNEGKTKGKITMPFIANAEFAKVASYG
LLPNMILYIRDYRFQVAKGTNVLFMWSAASNFTPLLAGFLSDSYLGRFLTITSIASL
SFLGVLWLWLTAMLQVQKSPCDPTAAGSHGCSSTASQALAIYSAPALISIGSGGRP
CSLAFGADOLDNKENPKNVRLESFPGWYVASSAVAVLIAFTGIVYIQEHLGWKIGFG
VPVLMILAAFLFILLASPLYVTRGVTKSLFTGLAQAIYAAYKKRKLSDPHHDSFDCY
YHMKSEIKAPSKQLRFLNKAACLIANSREEEISDGFALPWRCLCTDKVEELKALIV
IPITSTGINMSINTSSSQLLQATSMRRLSRHSGSFQVPAGSGMFTIIALKWI
LYDRAVIPLASKIRGRPFRLSLRMLGLFMSFLMAISAMVESPRRKAISQGVAN
NSNVDDISAMWLVPQYVLHGLAEALTAIGQTEFFTEPPKSMSSIAASLFLGLGMAVA
SLLASVILNAVNELSTRNGKESWSDNIKNGHNYVYVWLAIMSEINVIYVICSWSY
GPLVDQVRNCRVNGVREEELIDVKGKFEKEDLSIVVKTN"
26156..26179
/rpt_family="(TA)n"
26779..26899
/rpt_family="(CAAAA)n"
join(<26899..26982,27574..>27987)
/genes="F915.5"
<26899..>27987
/genes="F915.5"
/notes="predicted by geneScan"
join(26899..26982,27574..27987)
/genes="F915.5"
/notes="hypothetical protein"
/codon_start=1
/protein_id="AAF29405.1"
/db_xref="GI:6850342"

translation="WMRYCQFTPYDPSSNMDDLKAFEDETWKYKKKEAWSLFDEVLY
SEEMKNMGYTLNIMVNECFKIGRDOAIQIFVKTALKKPNPDVTCYRNITTLNKO
GVLFPAEYFFEEMCSDRLYPDVFTYTTMIDAYLKAGKTEDALRISNKKWDAFLQVA
WLACL"
complement(27450..27491)
/rpt_family="(TA)n"
complement(join(<29019..29311,29402..29473,29582..>29780))
/genes="F915.6"
complement(<29019..>29780)
/genes="F915.6"
/notes="similar to PGP224 GB:AAD02554 from [Petunia x
hybrida]"
complement(join(29019..29311,29402..29473,29582..>29780))
/genes="F915.6"
/notes="unknown protein"
/codon_start=1
/protein_id="AAF29406.1"

```



```

/db_xref="GI:6850343"
/translation="MGRVTPSEEDSNGLVQOQTPNQRTVRPVSQFAPPNYQQAN
VNLVSGRWRWSTGLFDCQADQANAVLTTIVPCVTFQIAEVMDEGEMTCPLGTFMYLML
MPALCSHWVMSKRYREKMRKFNLEVEAPYSDCASHVLCPCCSLCOEYRELKIRNLDP
LGTNSKRQIQIYIYGPGLFDISQVW"
complement(30151..30268)
/rpt_family="(TA)n"
complement(31066..31652)
/gene="F915.7"
complement(31066..31652)
/gene="F915.7"
/note="A duplicate of this gene can be found 5kb away;

```

alignment_scores:

```

Quality: 677.50 Length: 1084
Ratio: 1.403 Gaps: 40
Percent Similarity: 44.557 Percent Identity: 23.432

```

alignment_block:

```
US-09-434-382-2 x AC022354/rev ..

```

```
Align seg 1/1 to reverse of: AC022354 from: 1 to: 46335

```

```

56 AsnThrValTyrLeuGlnValAlaAlaGly.....SerArgAspSe 70
||||| ..... |||..... |||..... |||..... |||.....
10410 ATCCCATCTTT...CAGTACTTGAACCTGGAATGATACAGGATAC 10364

70 rGlyAlaAlaLeuTyrValPheSerGluPheAsnArgTyrLeuPheAsn 87
||||| ..... |||..... |||..... |||..... |||.....
10363 ATCATCTCTGTATTACTCTCTTCCTTATGAACAAAGATTCATCTTTAATG 10314

87 ySgLyglu..... 89
||||| ..... |||..... |||..... |||..... |||.....
10313 CTGGAGAGGTAATAATAATCGCTTACAGATTTTTCAGCAGTTTAAAGTT 10264

90 .....GlyValG1 92
||||| ..... |||..... |||..... |||..... |||.....
10263 TGTACTTGCATCTACTTGTGTAATGTATGTGTTTTTAGGCGTTGCA 10214

92 nArgLeuMetGlnGluHisLeuLysLeuLysValAlaArg..... 104
||||| ..... |||..... |||..... |||..... |||.....
10213 ACGTTTTTGTACAGACATAAGATTAAACTATCAAAAGGTTTCTCT 10164

104 ..... 104
||||| ..... |||..... |||..... |||..... |||.....
10163 GTTTTCTCTTGTATGTGATGAGTGAGTGTATTAATTTTGTATGTTGG 10114

105 .....LeuAspAsnIlePheLeuThrArgMethIst 115
||||| ..... |||..... |||..... |||..... |||.....
10113 TGAATTTTTTTTTTACAGATAGATCAATGATTTTGTCTGCTGTTGCTC 10064

115 pSerAsnValGlyLeuSerGlyMetIle..... 125
||||| ..... |||..... |||..... |||..... |||.....
10063 AGAGACAGCAGGAGGACTTCCAGGTTTGTTCAGCTGTGACCTTTTCTCT 10014

125 ..... 125
||||| ..... |||..... |||..... |||..... |||.....
10013 TTGTTGCTTGCACCTTTTATGTTATGTTGGTGATGCTAAGTTTTTTTCAT 9964

126 .....LeuThrLeuLysGluThrGly..... 132
||||| ..... |||..... |||..... |||..... |||.....
9963 GTTTTCATTGAATAGCTCTTCTACTAATCTAGCTAGCTGGTATTGGTGAAGA 9914

133 .....LeuProLysCysv 137
||||| ..... |||..... |||..... |||..... |||.....
9913 AGGGCTTCGGTGAGTATTATTGACATAATTTTGTGTGTTTCTTCTGTG 9864

137 al.....LeuSer..G 140
||||| ..... |||..... |||..... |||..... |||.....
9863 TATGAAATTGCTCAAGTATTGTTTTTTTGTACAAAAGGTTGAATGTATGGG 9814

140 lyProGlnLeuGluLysTyrLeuGluAlaIleLysIlePheSerGly 156
||||| ..... |||..... |||..... |||..... |||.....

```

```

9813 GTCTTCGGATCTTAACATATTTAGTTGATCAATGAAGTCGTTTATTCCA 9764

157 ProLeuLysGlyIleGluLeu...AlaValArgProHisSerAlaProG1 172
||||| ..... |||..... |||..... |||..... |||.....
9763 CGGCGAGCTATGTTTCATACACGAGTTTGGTCCCTCGAGTACTCTGA 9714

172 u.....TyrGluAspGluThrMetThrValTyrGlnIle..... 183
||||| ..... |||..... |||..... |||..... |||.....
9713 TCCTATTGTTCTTGTGAATGATGAAGTTGTGAAATATATCGCCCATTTACT 9664

184 .....ProIleHisSerGluGlnArgGlyLysHisGlnProTyrGln 198
||||| ..... |||..... |||..... |||..... |||.....
9663 TAAACCGTCTCATCTCGAGGA..... 9641

199 SerProGluArgProLeuSerArgLeuSerProGluArgSerSerAspSe 215
||||| ..... |||..... |||..... |||..... |||.....
9640 .....GACTC 9636

215 rGluSerAsnGluAsnGluProHisLeuProHisGlyValSerGlnArgA 232
||||| ..... |||..... |||..... |||..... |||.....
9635 T..... 9635

232 rGlyValArgAspSerSerLeuValAlaPheIleCysLysLeuHis 248
||||| ..... |||..... |||..... |||..... |||.....
9634 ..GGCAACAATCTGGGATCTCTCAGTTGTATATGATGTGTAATTACCT 9587

249 LeuLysArgGlyAsnPheLeuValLeuLysAlaLysGluMet...GlyLe 264
||||| ..... |||..... |||..... |||..... |||.....
9586 GAAATATTGGCAAAATTTGACTTTAGACAAGCTTAAGAAGCTTTTGTGT 9537

264 uProValGlyThrAlaAlaIleAlaProIleAlaAlaValLysAspG 281
||||| ..... |||..... |||..... |||..... |||.....
9536 TAAGCCTGCT.....CCAAATATAGTAGACTTTCAGTCCG 9502

281 lyLysSerIleThrHisGluGlyArgGluIleLeuAla..... 293
||||| ..... |||..... |||..... |||..... |||.....
9501 GTGAATCAGTGAAGTCAGATGAAGAGATATTACGGTATACCTGCCGATTA 9452

293 ..... 293
||||| ..... |||..... |||..... |||..... |||.....
9451 CCAATGACTGTTTTTTTCTCTCTCTCAGTGGAGATATTCTATGAAATTG 9402

294 .....GluGluLeuCyst 298
||||| ..... |||..... |||..... |||..... |||.....
9401 TAACGACTGTTTCATGCTTTTCATCTTAGTTTCATCCAAGTCATGTTATGG 9352

298 hrProAspProGlyAlaAlaPheValValGluCysProAspGlu 314
||||| ..... |||..... |||..... |||..... |||.....
9351 GACCTTCACCTCCCTGGTCCCATGTTCTCTCTGTGGACTGTCCACAGAG 9302

315 SerPheIleGlnProIleCysGluAsnAlaThrPheGlnArgTyrGlnG1 331
||||| ..... |||..... |||..... |||..... |||.....
9301 TCGCAGCTCGCAGAACTGTTCTCATTAACTTTCAGAGTATTATTATTC 9252

331 yLysAlaAspAlaPro.....ValAlaLeuValValH 342
||||| ..... |||..... |||..... |||..... |||.....
9251 ATCTCTGACGAGCAAAACCATTTGGAGCAAAAGTTTGTAAATTTGATTATAC 9202

342 isMetAlaProAlaSerValLeuValIleAspSerArgTyrGlnGlnTrpMet 358
||||| ..... |||..... |||..... |||..... |||.....
9201 ATTTGAGCCCATCTCTCTGTGACAGTAGTCTCTACTTACCAAGCTGGATG 9152

359 GluArgPheGlyProAspThrGlnHisLeuValLeuAsnGluAsn..... 373
||||| ..... |||..... |||..... |||..... |||.....
9151 AAGAAGTTTCATTG...ACTCAACACATTTCTTCTGCTGCCCATCAAGGTT 9105

374 .....CysA 375
||||| ..... |||..... |||..... |||..... |||.....
9104 TCTCCCTCTCTTATAATTTGTTTCTCATCAAAACCTGTTCCGGTAAATTT 9055

375 laSerValHisAsnLeuArg..... 381
||||| ..... |||..... |||..... |||..... |||.....
9054 GCTCCGCTGCTTATGTTGCGTTTGTGTATATTATTTCAGGAAGAAATATGCG 9005

```


clone:MYA6.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi
 1 (sites)
 Sato,S., Nakamura,Y., Kaneko,T., Kato,T., Asamizu,E. and Tabata,S.
 Structural analysis of Arabidopsis thaliana chromosome 3. I.
 Sequence features of the regions of 4,504,864 bp covered by sixty
 P1 and TAC clones
 DNA Res. 7 (2), 131-135 (2000)
 20277480
 2 (bases 1 to 75289)
 Sato,S., Nakamura,Y., Kaneko,T., Kato,T., Asamizu,E. and Tabata,S.
 Direct Submission
 Submitted (01-FEB-1999) to the DDBJ/EMBL/GenBank databases.
 Yasukazu Nakamura, Kazusa DNA Research Institute, Department of
 Plant Gene Research: 1532-3, Yana, Kisarazu, Chiba 252-0812, Japan
 (E-mail:ynakamukazusa.or.jp, Tel:81-438-52-3935,
 Fax:81-438-52-3934)
 Address for correspondence: kaos@kazusa.or.jp
 For the latest information on annotation of this clone, please see
 http://www.kazusa.or.jp/kaos/cgi-bin/agd_graph.cgi?c=MYA6
 Genes with similarity to proteins in the databases are described in
 'product' or 'note' qualifiers. Genes that have no significant
 protein similarity are described as 'unknown protein'.
 The software programs used to predict genes include: Grail
 (Informatics Group, Oak Ridge National Laboratory,
 http://compbio.ornl.gov/Grail-1.3/),
 GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html),
 NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of
 Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and
 SplicePredictor (Volker Brendel, Stanford University,
 http://gremli1.zool.iastate.edu/cgi-bin/sp.cgi).
 Genes encoding tRNAs are predicted by tRNAscan-SE
 (Sean Eddy, Washington University School of Medicine, St. Louis,
 http://genome.wustl.edu/eddy/tRNAscan-SE/).
 This sequence may not be the entire insert of this clone. It may be
 shorter because we remove overlaps between neighboring submissions.
 The 5' clone is MS11 and the 3' clone is MDC8.
 Location/Qualifiers
 1..75289
 /organism="Arabidopsis thaliana"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /chromosome="3"
 /clone="MYA6"
 /clone_lib="Mitsui pi"
 complement(join(483..4904,4991..5106,5192..5636,
 6082..6469))
 /note="gene_id:MYA6.1
 unknown protein"
 /codon_start=1
 /evidence=not_experimental
 /protein_id="BAB01260.1"
 /db_xref="GI:9279703"
 /translation="MERGMNQGSLHWSQSKPTPTMLDRALSSRRRHSADLSASGE
 SGTDESKRPHIVLLASNFLSRIGHOWPCLILALLFLVLLISVAFHSHSEVCIS
 RFDPAARIGFGLDSEDCALGPWCRSKHGEVENTSKDLKGLGEFPIYETRP
 IKNNYMGFDHSGNFMFARWLAPDMIESGAPKGSHTWVLRQAMPDTPMISLTPRH
 PEKYLKRGPDYNGCITYFAGKEDVFGSDVKMVLRRKHGIDLSRIVVDFEDHQRNL
 KRLQALKAGPRHLIFEDNYDTGCDYSLRQICDSHIRGGGHSFCFSDHMRKSK
 RKKFWKAVDTEELCGPGTWWGVGEMRDDFNITNTPISYNQHFQNSRYVESILDVY
 WELPPVAGPSLTHQSRVDPARATPPIVADGKHLRFQRLGRLDKSVFNGYQTQVYLE
 ISKPSG"
 complement(7397..8479)
 /note="gb|AAF25964.1
 gene_id:MYA6.2
 similar to unknown protein"
 /codon_start=1
 /evidence=not_experimental
 /protein_id="BAB01261.1"
 /db_xref="GI:9279704"
 /translation="MSKFLPEBLAIEILVRLSMKDLARFCVCKTWRDLINDPGTET
 YRDMSPKPVSYDKNFYMLDVECKHPVITNKLDPDQSDMIDESTCVLHCDGLCVT
 LKNITLWMPNDFQKFIKVPNGIYQDSNIGLFGDVPDHDYKVVFIDRLDVSIAHV
 FEFTSGESLIRSYPDHVRDRRTFDOYLYWIAVRSADREILCNLSMSSFSIS
 LPLVYNGVTSYMLGVTSQKLCITEYEMCKEIRISVMEKTSKWSKILISLSSFSIS
 VQDRIYQVFEVSEFTRANDLVVITFTGYNDHFEMPEERTKKMKFLYKTGERSSEVR
 FCNPLAGURFLCECVETUKIVNRFI"
 join(10577..10657,10791..11094,11019..11131,11203..11292,
 11366..11429,11555..11619,11726..11932)
 /note="gene_id:MYA6.3
 unknown protein"
 /codon_start=1
 /evidence=not_experimental
 /protein_id="BAB01262.1"
 /db_xref="GI:9279705"
 /translation="MENLNHKWDRCSSSHQTPLSFGKNOKIOGQOREVTFHFVSLPL
 AYPDLTKNIEAFONSVLNGDKPLKFQSTLAEMGIEKSIETVPTKTLTVVMMLKE
 NNEVYKQNILOVLSFSEKQSIQSVRQALKNRVFIIRLGLSCMSGLDKTRVLYA
 PVEEVGHVLIIDAFENVGAGDAKSLRKLHATLMNASYRKDKSKMDTFDAREIHKEF
 ENKDMGTVLIIEAHISORYKYDPNGYFHCASLPPPHK"
 join(13499..13671,13836..13881,14227..14264,14340..14419,
 14536..14623,14732..15011,15103..15215,15295..15360,
 15467..15572,15667..15731,15855..15895,15984..16167)
 /note="gene_id:MYA6.4
 unknown protein"
 /codon_start=1
 /evidence=not_experimental
 /protein_id="BAB01263.1"
 /db_xref="GI:9279706"
 /translation="MIFYQSNKTNQFIFFLAWGLEKSFDFVLFLFMSHVVCVGLT
 VLGRLVDVFNRRGSEKAMDYKKQIKMVGKSLIKFIRGEGTQMKLEEMGVKII
 LPSSRNKDIHIEGGSVDCVTKASKRIATIIDEVVSPSLDYSHEVSLPLAHPLEVL
 KLVNFQNSILGTHIASDQDQANRATTSVAVDLKANSENOVNVKTSIPISVYPP
 KAKSKSTLLDLGIEKTSIFIKPSTFHLTVMLKLNKDRVNAACDLKSIPISSVNDAL
 DNKPVFIRLGLDCMRGLDKTRVLYAPVEELGDEGLLRACRILDFITDAFVKAGL
 VLEQDAKQSLKLVHTVMNRRKNNKMMETFDAREIHKQFGNEDWGEYLIQEAH
 LSQRFVQNGYRCCGISPPFGEQRA"
 join(18398..18524,18628..18878,19278..19652)
 /note="gb|AAC49281.1
 gene_id:MYA6.5"
 /codon_start=1
 /evidence=not_experimental
 /product="delta tonoplast intrinsic protein"
 /protein_id="BAB01264.1"
 /db_xref="GI:9279707"
 /translation="MAGVAFGSFDDSFSLASIRAYLAEFISTLLEFVAGVSAIAYAK
 LITSDALDTPGLVAIVACHGFALFVAVAGISGHNVPATVFLGAVGGQITVITGV
 FYTAQLLGSTAAECFLKLYVTGGLAVPHTSVAAGLGSIEGVVMEIITIFALVYTVAT
 AADPKKGLGTIAPLATGLIVGANILAAAGPSGSGMNPARSFGPAAGDFSGHVVYV
 VGPLTGGGLAGLIYGNVPMGSEHVPLASADF"
 complement(join(19955..20047,20150..20212,20303..20365,
 20457..20513,20611..20721,20957..21184))
 /note="gene_id:MYA6.6
 unknown protein"
 /codon_start=1
 /evidence=not_experimental
 /protein_id="BAB01265.1"
 /db_xref="GI:9279708"
 /translation="MGVQLSGSLVASLPPNHSHTKLNKPNYSYFRSKHNAART
 KTVRAISTAPASQPPADEPDPVAFVHSVLDPDTPDVRHRRANGGQKLRDIM
 LDSNIELYSPYSKPLSNCAGVTCATVAFVHSVLDPDTPDVRHRRANGGQKLRDIM
 ACQTNVGPNDSGLVVIQQLPEWKAHWNIPKINPDDDDLETST"
 join(22261..22656,22735..22824,23038..23092,23187..23230,
 23347..23742,23833..24098,24182..24326,24415..24465,
 24561..25553,25635..25982)
 /note="gb|AAF29402.1
 gene_id:MYA6.7
 strong similarity to unknown protein"
 /codon_start=1
 /evidence=not_experimental
 /protein_id="BAB01266.1"
 /db_xref="GI:9279709"
 /translation="MLTSSMPQNLSLFGFSPLKSSSFALILRPFSLYPPPIFASSSPAP

2315	tGluSerAsnGluAsnGluProHisLeuProHisGlyValSerGlnArgA	232
	: : : : :	
23577	TGGCAGCAAAAGCTGGTGA...	23595
232	rgGlyValArgAspSerLeuValValAlaPheIleCysLysLeuHis	248
	: : : : : : : : : :	
23596ACAGCGTTATATATGTATGTGAATTACT	23625
249	LeuTysArgGlyAsnPheLeuValLeuLysAlaLysGluMetGlyLeuPr	265
	: : : : : : : : : :	
23626	GAATCAAAGGAAAGTTTGATCCAAAGAGGCCATGGCTCTTGATTGAG	23675
265	ovalGlyThrAlaAlaIleAlaProIleIleAlaValLysAspGlyL	282
	: : : : : : : :	
23676	AGCTGGT.....CCAAAGTACAGTTATCTTCAGTCTGGAC	23710
282	ysSerIleThrHisGlyArgGluIleLeuAla.....	293
	: : : : : : : : : : : : :	
23711	AGTCAGTCAAGTCAGATTTTAAAGATATTACGGTATGTCTCAAAGCAA	23760
293	293
23761	TCATTGGGAATCTTCACAACCCAAATATCAGTTCTATACAAGCAGACTAA	23810
294GluGluLeuCysThrProp	300
	: : : : : : : : : : : : : : : :	
23811	CAGCTTCACGGTTAATCATAGTTTCATCCGAGTGATGTGATGGACCTT	23860
300	roAspProGlyAlaAlaPheValValValGluCysProAspGluSerPhe	316
	: : : : : : : : : :	
23861	CAGTTCCTGGCTGTTTGTGTGGATTGTCCAACAGACTCAT	23910
317	IleGlnProIleCysGluAsnAlaThrPheGlnArgTyrGlnGlyLysAl	333
	: : : : : : : : : : : : : : : : : : : : :	
23911	GCCGAGGAACTGCTATCTATCCCGTCTATGAAGACTTATTATTCTGTCT	23960
333	aspAlaPro.....ValAlaLeuValValHisMeta	344
	: : : : : : : : : :	
23961	TGACAATTCAACAGATGGAGCTAGCTTGTTAATTGTATTATTCATCTAA	24010
344	laProAlaSerValLeuValValAspSerArgTyrGlnGlnTrpMetGluArg	360
	: : : : : :	
24011	GTCCAGCTTCTGTAAACAATAGTTCTACATACCGAAGCTGATGAAGAGA	24060
361	PheGlyProAspThrGlnHisLeu.....Val..	369
	: : : : : : : : : :	
24061	PTC...CATTTGCCCCACACATTCGCTGGACATGAAGCGTGAGTCTT	24107
370	LeuAsnGluAsp.....	373
	: : : : : : : : : :	
24108	CTCAACTATGGTTCTTCTCATAAAGAACTGTGAATTTATTGTCTGA	24157
374CysAlaSerValHisAsnLeuArg.....	381
	: : : : : : : : : :	
24158	TATGTAGTTTGGTGTGTCCAGGAAGAATATGGAGTTTCCAATCTGA	24207
382SerHisLysIleGlnThrGlnLeuAsnLeuIleHisProAspIle	396
	: : : : : : : : : : : : :	
24208	GAGCCAGTCTAGAAATCACGCCAGACTTAACTATTATGTCACAGTTC	24257
397	PhePro.....	398
	: : : : : : : : : :	
24258	TTCCAGCTCTGGATTTTGGTCTCATCAACATGATATACTCGATAAA	24307
398	398
24308	CCCCACTCTTTTGACAAAGTTTACAACGCCTTTTCTGATATTTTCTTTT	24357
399Leu	399
	: : : : : : : : : :	
24358	GAGTTGAATCTGGCTGTGTTCATTTAGTCAAGTTGTGTCTTTTCTCTC	24407

[illegible]

```
25302 ACAGACAATCTTCA...GCTTTACCATCTTGAAGAACTG...AAAA 25345
627 rLeuLeuArgThrCysAspLeuGluGluPheGlnThrCysLeuValArgH 644
25346 AGTCCTCGCGGNAATGGGTGGAGCATTTGATTAGCTTCTCTGTGTGC 25395
644 IsCysLysHisAlaPheGlyCysAlaLeuVal..... 654
25396 ATTGCGCTCAAGCCCTTGGGTTTGAAGAGCAGCAGAGAAAAAAC 25445
655 .....HisThrSerGlyTrpLysValValTyrSerGlyAspTh 667
25446 ATTGCGGAGATGAGATCCGGGGTGGAAATGGTTTACTCGGAGACAC 25495
667 rMetProCysGluAlaLeuArgMetGlyLysAspAlaThrLeuLeuI 684
25496 AAGGCTTGTCTGAGATGGTTGAAGCTTCAAAAGGACACAGTTCTTA 25545
684 leHisGlu..... 686
25546 TACATCAGGTGCTTCTTCTATATAAGCAAAACGTCTGTAACACATTTACATC 25595
687 .....AlaThrLeuGl 690
25596 TCATTTCTTAAATTTACCATCTTTTCAAACTTATTAACAGCGCGCTTGA 25645
690 uAspGlyLeuGluGluAlaValAlaGluLysThrHisSerThrSerG 707
25646 GGATCGGTGGTAGAAGAGCTGTGGCTTAAACCATAGCAGCACTAAG 25695
707 InAlaIleSerValGlyMetArgMetAsnAlaGluPheIleMetLeuAsn 723
25696 AAGCAATCAAGCTTGCATCTTCAGCAGGTGTATACCGAACCGTACTGACA 25745
724 HisPheSerGlnArgTyrAlaLysValProLeuPheSerProAsnPhse 740
25746 CATTTAGCCAGAGATACCCAAAGATCCCGGTGATTGACGAATCACATAT 25795
740 rGluLysValGlyValAlaPheAspHisMetLysValCysPheGlyAsp 757
25796 GCACATACATGCTGCTTGTATGATGATGATGATGATGATGATGATG 25845
757 heProThrMetProLysLeuIleProProLeuLysAlaLeuPheAlaGly 773
25846 TACATGCTCTTCCAAAGATTTTACCTTACTTCAAACTCTATTTCAGAAAC 25895
774 AspIleGluGluMetGluGluArgGluLysArgGlu 786
25896 CAAGTGGTTGAAGAAGAAGAAGAAGAAGAAGAGACTGAT 25934
seq_name: gb_pl3:SCYKR079C
```

```
seq_documentation_block: 3685 bp DNA PLN 11-AUG-1997
LOCUS SCYKR079C
DEFINITION S.cerevisiae chromosome XI reading frame ORF YKR079c.
ACCESSION Z28304 V13137
VERSION Z28304.1 GI:486556
KEYWORDS
SOURCE
  baker's yeast.
  Saccharomyces cerevisiae
  Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
  Saccharomycetaceae; Saccharomyces.
REFERENCE
  1 (bases 1 to 3685)
  Pohl,T.M. and Pohl,F.M.
  Unpublished
REFERENCE
  2 (bases 1 to 3685)
  MIPS.
  Direct Submission
  Submitted (09-MAY-1994) Data collected by MIPS on behalf of the
  European yeast chromosome XI sequencing project. MIPS at the
  Max-Planck-Institut fuer Biochemie, Am Klopferspit 18a D-82152
  Martinsried, FRG; E-mail: Mewes@hpmc.mips.biochem.mpg.de
```

```
FEATURES
    source
        Location/Qualifiers
            1..3685
                /organism="Saccharomyces cerevisiae"
                /db_xref="taxon:4932"
                /chromosome="XI"
                complement(80..2596)
                /note="ORF YKR079c"
                /codon_start=1
                /protein_id="CAA82158.1"
                /db_xref="GI:486557"
                /db_xref="SWISS-PROT:P36159"
                /translation="MTFTPTSDTKHPLLVQSAHGEKYFFKIGGSGRSRLTE
                NKIRISKLKIDIFLTGELNMSDIGLPGMILTIAQQKSNLVLHYGNDILNYIVTWRY
                FVERFGDLNHIKMKVEYKDKIIAIVKSENVKNGEEDRLGVDSPQKGLRISIVAK
                MPFHAPTDRIYDPSDPHLNVELDLDAKVEVSTNYEISPSVGRGFKVEAEIKLGPV
                KGPLEAKLTGQITLDNGIVTPEOVLENERHFAKVLILIDIPDLYLNFAVEKFKDY
                DCAELGMVYFELGDEVTINDNLEAFIDIEKNNYKVNHMISHNKISPNWISPEFGSAL
                TTUKALKAVNNLPLTKDRVSKDFDFTPLSGTSSCKSOEPLNTIIBKDNH
                IFSQNTVTTFEPERMEEPKMCNGEADFSWQEIFEEHVKPEFLPADVDVFINNQ
                LHVDFNNSAEKKKHVEIITLTGTSALPSKRVNVSTLVKVPFTDADGNTINRMILD
                AGENTLGTTHRMFSQIAVKSIFODLKMVYLSHLADHHLGIIISLVNEWYKYNKDDETS
                YIVVTPWQYHKFVNEMLVLENKILKRYISCEHFINDSFVRMOTOSVPLAEFNEI
                LKENSQESNRKLELDRDSSYRDVLIHQWEDLSIEYFQTCRAIHCDAWYSNITFR
                MDENNEHTFKVYSIGDTRPNIEKFSLEIGYNSDLLIHEATLENLLEDAVKKKHCTI
                NEAIGVSNKMNARKLILTHFSQRYPKLPQLDNNIDYWAREFCFAFDSMIVDYEKIGSQ
                ORIFPLLNKAFVEEKEDVDVDSVQDLVLEVKLKHKN"
BASE COUNT      1107 a      716 c      629 g      1233 t
ORIGIN

alignment_scores:
    Quality: 613.00      Length: 914
    Ratio: 1.288      Gaps: 38
    Percent Similarity: 52.079      Percent Identity: 25.383

alignment_block:
    US-09-434-382-2 x SCYKR079C/rev ..
    Align seg 1/1 to reverse of: SCYKR079C from: 1 to: 3685

29 ArgGluArgProArgLysAspProLeuArgHisLeuArgThrArgGluLys 45
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2671 AGAAGGACCTTAAAGAACACAGACTAATAGACGACACCCGCAAGA 2622

45 sArgGlyProSerGlyCysSerGlyGlyProAsnThrValTyrLeuGlnV 62
||||| ||||| ||||| ||||| ||||| ||||| |||||
2621 AAAGAGTCTCGAAGGTTTCA.....ACGATGTTACATTATACCA 2578

62 alValAlaAlaGlySerArgAspSerGlyAlaAlaLeuTyrValPheSer 78
||||| ||||| ||||| ||||| ||||| ||||| |||||
2577 TCACCCATCTACATCGGATACAAAGCACCACCTTGTCTAGTCCAGTCT 2528

79 GluPhe....AsnArgTyrLeuPhe....AsnCysGlyGluGlyValGlnAr 93
||||| ||||| ||||| ||||| ||||| ||||| |||||
2527 GCACATGGGGAAGATATTCTTCGGTAAATTTGGTGAAGGATCCCAAG 2478

93 gLeuMetGlnGluHisLysLeuLysValAlaArgLeuAspAsnIlePheL 110
||||| ||||| ||||| ||||| ||||| ||||| |||||
2477 GAGTCTGACTGAAATAAGATCAGGATATCCAAATTTGAAGGATATTTCC 2428

110 euThr...ArgMetHisTrpSerAsnValGlyGlyLeuSerGlyMetIle 125
||||| ||||| ||||| ||||| ||||| ||||| |||||
2427 TTACTGGTGAATTAACCTGGTTCAGATATAGTGGATTACCTGGAATGAT 2378

126 LeuThrLeuLysGluThrGlyLeuProLysCysValLeuSerGlyProPr 142
||||| ||||| ||||| ||||| ||||| ||||| |||||
2377 TTGACTATTGCTCAAGGGAAGAAATCTTGTTTTGCTATTACGGCAA 2328

142 ogLLeuGluLysTyrLeuGluAlaIleLysIlePheSerGlyProLeuL 159
||||| ||||| ||||| ||||| ||||| ||||| |||||
2327 TCACATTTTGAATTACATAGTTTCCACTTGGAGATACCTCGTCTTTAGAT 2278

159 ysGlyIleGluLeuAlaValArgProHisSerAlaProGlu..... 172
```

```

2277 TCGGATAGACTTG.....AGGATCACATTATGAAGACAGGAAGTA 2234
|||||.....|
173 TyrGluAspGluThrMetThrValTyrGlnIleProIle.....Hi 186
|||||.....|
2233 TATAAGATAAGATAATAGCTGTAAATCCTTTAATGTTCTGAANAATGG 2184
|||||.....|
186 sSerGluGlnArgArgGlyLysHisGlnProTrpGlnSer..... 199
|||||.....|
2183 GGGGGAAGACAGGTTAGGCGCTTCGATAGTTTCAAAAAGGTGTATTAC 2134
|||||.....|
200 .....ProGluArgProLeuSerArgLeu 207
|||||.....|
2133 GTTCCATAGTAGCAAAAATGTTCCCAACATGACCCACCGATAGGTAC 2084
|||||.....|
208 SerProGluArgSerSerAspSerGluSerAsnGluAsnGluProHisLe 224
|||||.....|
2083 GATCCT.....TCTAGTAGTCGGCACTGGAATGTAGAGTTGCCTGACTT 2040
|||||.....|
224 uProHisGlyValSerGlnArgArgGlyValArgAspSerSerLeuValV 241
|||||.....|
2039 A.....GACGCCAAAGTGAAG 2023
|||||.....|
241 alAlaPheIleCysLysLeuHisLeuLys.....ArgGlyAsnPheLeu 255
|||||.....|
2022 TTCTACGAATACGAGATTTCATTCAGTCAGTGAGGGGTAAATTTAA 1973
|||||.....|
256 ValLeuLysAlaLysGluMetGlyLeuProValGlyThrAlaAlaIleAl 272
|||||.....|
1972 GTGGAGGAAGCTATTAACTAGGTGTCGGAAGGT..... 1937
|||||.....|
272 aProIleIleAlaValLysAspGlyLysSerIleThr.....HisGluG 288
|||||.....|
1936 .CCCTATTGCAAAAGTTAAACCAAGGCCCAACAATTACTTTGGATAACG 1888
|||||.....|
288 lYargGluIleLeuAlaGluGluLeuLysThrProProAspProGlyAla 304
|||||.....|
1887 GTATTGTTGTAACTCGGGAACAGGTATTTGGAGAAATGAACGTCAATTTGCC 1838
|||||.....|
305 AlaPheValValValGluCysProAspGluSerPheIleGlnProIleCy 321
|||||.....|
1837 AAAGTATTGATCCTGGATATCCAGATGACCTATATTG..... 1799
|||||.....|
321 sGluAsnAlaThrPheGluArgTyrGlnGlyLysAlaAspAlaProValA 338
|||||.....|
1798 ...AACGCTTTCGTAGAAAAATTCAGAGATTATGATTGCTGAGCCTTG 1753
|||||.....|
338 lAlaValValHisMetAlaProAlaSerValLeuValAspSerArgTyr 354
|||||.....|
1752 GCATGTGTATTATTTCTGTGTGATGAGGTACCATTAAATGATAATCTA 1703
|||||.....|
355 GlnGlnTrpMetGluArgPheGlyProAspThrGlnHisLeuValLeuAs 371
|||||.....|
1702 TTCGCGCTTCATGACATATTGAG.....AA 1677
|||||.....|
371 nGluAsnCysAlaSerValHisAsnLeuArgSerHisLysIleGlnThrG 388
|||||.....|
1676 AAACAATTTATGTTAAAGTAAATCATATGATATCCAC..... 1640
|||||.....|
388 lLeuAsnLeuIleHisProAspIlePheProLeu.....Leu 400
|||||.....|
1639 ....AATAAAATTTCTCCAAACACGATATCATATTTTCGGTTCTGCATTG 1595
|||||.....|
401 ThrSerPheArgCysLys.....Ly 407
|||||.....|
1594 ACCACATTGAATTAAGGCACACTACAGTAATAATTAATTAATTAACAAA 1545
|||||.....|
407 sGluGlyProThrLeuSer.....ValProM 416
|||||.....|
1544 AACAGATCGTGTGTTTTCGAAGGACTTCAGACAGATTCGATACACCAC 1495
|||||.....|
416 etValGlnGlyGluCysLeuLeuLysTyrGlnLeuArgProArgArgGlu 432
|||||.....|

```

```

1494 TCAGCAGAGGTACATCTATGTGTAATCCAGGAAGAGCCTTTGAATACA 1445
|||||.....|
433 .....TrpGlnArgAspAlaIleIleTh 440
|||||.....|
1444 ATAATAGAGAAGGATAACATTCATATTTTTCACAAAACAGACAGTAAC 1395
|||||.....|
440 rCysAsnPro..... 443
|||||.....|
1394 TTTCTGAACCATTTTCGGATGAACGAAGAACCGATGAATGCAACATCAACG 1345
|||||.....|
444 .....GluGluPheIleVal 448
|||||.....|
1344 GTCAAGTGGCGGATTTCTCGTGCAAGAAATTTTCGAAGAACAT...GTA 1298
|||||.....|
449 GluAlaLeuGlnLeuPro.....AsnPheGl 457
|||||.....|
1297 AAACCATTAGAATTTCCCTTAGCTGATGCGATACACTTATCAATATCA 1248
|||||.....|
457 nGlnSerValGlnGluTyrArgArgSerAlaGlnAspGlyProAlaProA 474
|||||.....|
1247 ACTACACGCTGGATAACTTTAAACAATTCAGCAGAA..... 1214
|||||.....|
474 laGluLysArgSerGlnTyrProGluIleIlePheLeuGlyThrGlySer 490
|||||.....|
1213 .....ANGAAGAAACACGTTGAATTTATCACCTTAGGAACCGGTAGT 1172
|||||.....|
491 AlaIleProMetLysIleArgAsnValSerAlaThrLeuValAsnIle... 506
|||||.....|
1171 GCATTGCCITCTAAATATAGAAACGTTGTCTCCACACTTGTAAAGTTCC 1122
|||||.....|
507 .....SerProAspThrSerLeuLeuLeuAspC 516
|||||.....|
1121 TTTTACTGACGCGCATGGAATACCATAAATAGAAACATTATGCTAGATG 1072
|||||.....|
516 yscGlyGluGlyThrPheGlyGlnLeuLysCysArgHisTyrGlyAsp...Gln 531
|||||.....|
1071 CTGGTGAATAACTTTAGGTACCATACACAGGATGTTTTCTCAGCTAGCA 1022
|||||.....|
532 ValAspArgValLeuGlyThrLeuAlaAlaValPheValSerHisLeuHi 548
|||||.....|
1021 GTCAAGTCAATATTTCCAGGATTTGAAAATGATATATCTGACTCAGTCCA 972
|||||.....|
548 sAlaAspHisHisThrGlyLeuProSerIleLeuLeuGlnArgGluArgA 565
|||||.....|
971 TGCAGACACCATTTGGGAATAATACGCGTGCTA.....AATCAATGGT 928
|||||.....|
565 lAlaLeuAlaSerLeuGlyLysProLeuHisProLeuLeuValAlaPro 581
|||||.....|
927 ACAAATATAACAAGGATGATGAACGAGTTATATATATGTTGGTTACTCCA 878
|||||.....|
582 AsnGlnLeuLysAlaTrpLeuGlnGlnTyrHisAsn..... 593
|||||.....|
877 .....TGG.....CAATATCACAAATTTGTTAATGAATG 849
|||||.....|
594 .....GlnCysGlnGluValLeuHisHisIleSerMetIleProA 607
|||||.....|
848 GTTAGTTCTAGAAAATAAAGAGATTTTAAAGAGAATCAAAATACATAAGTT 799
|||||.....|
607 la..... 607
|||||.....|
798 GTGAGCATTTTCATCAATGATTCGTTTGTAGAAATGCGACACACAATCTGTT 749
|||||.....|
608 .....LysCysLeuGlnGluGlyAla..... 614
|||||.....|
748 CCTTTGCGCAGAGTTCAATGAAATATTGAAGAAATAGCAATCAAGAAATC 699
|||||.....|
615 .....GluIleSerSerProAlaValGluArg.....L 624
|||||.....|
698 AAACAGAAAACTGGAACCTGGATAGATGATCTTCATATAGGGATGTTGACT 649
|||||.....|
624 euIleSerSerLeuLeuArgThrCysAspLeuGluGluPheGlnThrCys 640
|||||.....|
648 TGATCAGACAAATGTATGAGGATTTATCGATAGATATTTTCAAACTTGC 599
|||||.....|

```

641 LeuValArgHisCysLysHisAlaPheGlyCysAlaLeuVal..... 654
:: ||||| |||||
598 AGAGCTATACATTGTGACTGGGCATATTCGAACCAATTACCTCCGAAT 549

655HisThrSerGlyTrpLysValValThrSerGlyA 666
|||::: ||::: ||||| |||||
548 GGACGAAACAATGAGCATATATCA...TTCGAAGTTTCATATTACGCG 502
666 spThrMetProCys.....GluAlaLeuValArgMetGlyLysAspAla 680
||||| ||| :::: ||::: |||||
501 ATACAAGACCTAATCATCCAGAAATTTCCCTCGAAATAGGCTAATAATCA 452
681 ThrLeuLeuLeHisGluAlaThrLeuGluAspGlyLeuGluGluAla 697
|||::: |||||
451 GATCTATTAAATTCACGAAGCTACACTAGAAATCAGCTACTGAGGAGTGC 402
697 aValGluLysThrHisSerThrThrSerGlnAlaIleSerValGlyMetA 714
||||::: |||| ||| :::: |||||::: |||||::: ;
401 CGTGAAGAAAACACATCGCACTATTAAATGAAGCAATCGTGTTCGAACA 352
714 rgMetAsnAlaGluPheIleMetLeuAsnHisPheSerGlnArgTrpAla 730
::: ||::: ||::: |||||
351 AAATGAATGCTAGGAAGTTGATCTTAACACACTTTTCCAGAGATATCCC 302
731 LysValProLeuPheSerProAsn.....PheSerGluLysValGl 744
|||::: ||||| ||| :::: |||||
301 AAATTGCCCCAATTAGACAATAATATTGATGTGATGGCAGAGAAATTTTG 252
744 yValAlaPheAspHisMetLysValCysPheGlyAspPheProThrMetp 761
||||||| |||| ||| ::: |||
251 TTTTGTGCTTCGACAGTATGATCGTTGATTATGAGAAATTCGTGAACAGC 202
761 rOLysLeuIleProLeuLysAlaLeuPheAlaGlyAspIleGluGlu 777
::: ||||| |||||
201 AGCGTATTTTCCACTCTCGAATAGGCCATTT.....GTTCGAAGAA 161

778 MetGluGluArgArgGluLysArgGluLeuArgGlnValArg 791
|||||::: |||||
160 AAGGAAGAAGAAGAAGATGTTGATGACGCTTGAAGACGCTACAA 119

seq_name: qb_pl1:AF188714

seq_documentation_block:	
LOCUS	AF188714 19158 bp DNA 31-JAN-2000
DEFINITION	Americella nidulans chromosome VIII cosmid SW06H01, genomic sequence.
ACCESSION	AF188714
VERSION	AF188714.1
KEYWORDS	GI:6707311
SOURCE	.
ORGANISM	Aspergillus nidulans. Aspergillus nidulans. Eukaryota; Fungi; Ascomycota; Eurotiales; Trichocomaceae; Emericella.
REFERENCE	1 (bases 1 to 19158)
AUTHORS	Akano,T. and Akita,O.
TITLE	Direct Submission
JOURNAL	Submitted (22-SEP-1999) National Research Institute of Brewing, Kagamiyama 3-7-1, Higashi-Hiroshima, Hiroshima 739-0046, Japan
REFERENCE	2 (bases 1 to 19158)
AUTHORS	Gomi,K.
TITLE	Direct Submission
JOURNAL	Submitted (22-SEP-1999) Department of Agricultural Science, Tohoku University, 1-1 Tsutsumidori-Anamiamachi, Aoba-ku, Sendai 981-8555, Japan
REFERENCE	3 (bases 1 to 19158)
AUTHORS	Kunihiro,S., Nakagawa,M. and Machida,M.
TITLE	Direct Submission
JOURNAL	Submitted (22-SEP-1999) Molecular Biology Department, National Institute of Bioscience and Human-Technology, Higashi 1-1, Tsukuba Ibaraki 305-8566, Japan
FEATURES	Location/Qualifiers
source	1..19158

```

organism="Aspergillus nidulans"
/dbx_ref="taxon:5072"
/chromosome="viii"
/clone="cosmid SW06H01"
/notes="overlaps cosmid SW05E08 deposited in GenBank"
Accession Number AC000133
4815 a 4709 c 4726 g 4908 t
BASE COUNT
ORIGIN

```

```

alignment_scores:
  Quality: 610.00      Length: 1183
  Ratio: 1.180         Gaps: 42
  Percent Similarity: 43.702  Percent Identity: 22.316

alignment_block:
  US-09-434-382-2 x AF188714/rev ..

  Align seg 1/1 to reverse of: AF188714 from: 1 to: 19158

      7 LeuLeuArgSerAlaAlaGlyArgThrMetSerGln.....G1 19
      18045 TTACTTCACCGTCGCCCATCCGCCACTGTTACAAATAAAGAGAGGCCT 17996

      19 yArgThrIleSerGlnAlaProAla.....ArgArgGluArgp 32
      17995 TCGCGCATACTCTACAGCTCCCGAGCCTCGATAGTCGGCTCCCTCAAGC 17946

      32 Ro...ArgLysAspProLeuArgHisIleuArgThrArgGluLysArgGly 47
      17945 CCTCCGAGAAAGACAATTCGAACATTATACGAAGAAGCAAAAGCGA... 17899

      48 ProSerGlyCysSerGlyGlyProAsnThr..... 57

      17898 .....TCTTCGCTCTACAATACATACACTATCTCGGAACACC 17861

      58 .....CCAAATCCAAAATTCGGAATTACGAAATGAAGTTCTACTATCAGATCT 17811

      17860 CCAATTCGAAATTCGGAATTACGAAATGAAGTTCTACTATCAGATCT 17811

      63 alAlaAlaGlySerArgAspSer...GlyAlaAlaLeuTyrrValPheSer 78
      17810 TGACGACCGCGAGCGGCTGATACCCAGCGCTCGCGGCTTCTCTGCATTTC 17761

      79 GluPheAsnArgTyrrLeuPhe...AsnCysGlyGluGlyValGlnArgLe 94
      17760 CCAGGGAACGTTATCTTTCCGCCCAATTCGACAGGGGGTTCCACAGAGC 17711

      94 uMetGlnGluHisLysLeuLysValAlaArgLeuAspAsnIlePheLeuT 111
      17710 ATGTACGGAACGAGGAGCAAAAGTTGACGGAGCTGCTCATGCTTTTCTGT 17661

      111 hr...ArgMetHisTrpSerAsnValGlyGlyLeuSerGlyMetIleLeu 126
      17660 CGGTCGCATGGATGGAGACAACTCGAGGCTTGATTTGGCATGATCTTT 17611

      127 ThrLeu..... 128
      17611 .....

      17610 ACGAAGCGCGATGCTCGCGCAGTTCAAAGAGGCTCTGGAAGCCCTCGA 17561

      129 .....LysGluPheGlyLeuProLys..... 135
      17560 GCGGAAAAGGAGCGAATTCACAGAGAGCGAGGTGGACCAAGAAAGATA 17511

      136 .....Cys 136
      17510 CGAAACAAACCGCCCAACGACGAGCTCGAAGATCCGCACACGACTTG 17461

      137 ValLeuSerGlyProProGlnLeuGluLysTyrrLeu..... 148
      17460 ACTGTCACGGTTCAAGAAATCTTGTCTATACCTTGGCTACTGGCGGAG 17411

      149 .....GluAlaIleLysIlePheSerGlyProLeuLysGlyI 161

```



```

17410 GYTTGCTTCAGACAGGTTTCCCGTGTCATACGAAGAGTAGCAGCCGG 17361
      ::::::::::: :::: ::::
161 leGluLeuAlaValArg.....ProHisSerAla 170
      :::::||||| ||| :::::
17360 AAAGTGTTCAAAGAGGCTCGGCACATGATCAAAGCGACCCCTTTGAAGTG 17311
      :::::||||| ||| :::::
171 proGluTyrGluaspGluThrMetThrValTyrGlnIleProIleHis 187
      ||| :::::||||| ::::: ||:::
17310 CCAGCTTTTCGGATGAAACATCAAGGCTCGACCATGCTATCAGCCC 17261
      :::::
187 rGluGlnArgGlyGlyHisGlnProTrpGlnSerProGluArg.... 202
      :::::
17260 GTCTTCTACTACACCACTTCG.....CAAGCCCTAAAGAGGA 17220
      :::::
202 ..... 202
17219 GTCTTCACGAATTTCCGGAAGATGTGCGGGTTTAGTGAGGTGGACCAG 17170
202 ..... 202
17169 CAATCCAAGGATCAATTTGTGGCGCAGCTGTGTTGTCTCATATGTTGA 17120
203 .....ProLeuSerArgLeuSerProGluArgSerSerAsp..... 214
      ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
17119 CTCAACTTGGACTCTGGATGCAATGGAGGAAACCCGCTTGGCCGAGCTCA 17070
215 .....SerGluSerAsnGluAsnGlu 221
      ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
17069 AAATGCCAGCGCAAACTTTGTACGAATCCCGAGACCAAGGACCTGGAC 17020
222 ProHis.....LeuProHisGlyValSerGlnArgArgGlyValAr 235
      :::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
17019 AAGTACACAGCGCCGCTTCTGGAGGAGATGAGGAG.....GTTC 16979
235 gaspSerSerLeuValVal..... 241
      ::::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
16978 TGAGATGACCGGTTCTGTCGTAAACCCCTGGCCCGGCTCGATCAGCA 16929
242 .....AlaPheIleCysLys 246
      ::::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
16928 AGATCCCAACAACACACCATCGTTGAGAGTCTATCTTACATCATCAAG 16879
247 LeuHisLeuLysArgGlyAsnPheLeuValLeuLysAlaLysGluMetGl 263
      ||| ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
16878 AACCCAGACCTCGCGGAAAGTTTGATGCTAAGAAGCAATCGCTCTGAA 16829
263 yLeuProValGlyThrAlaAlaIleAlaProIleIleAlaAlaValLysA 280
      :::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
16828 TGTCAACCCCTGGC.....CCTGACTTTGGTGTCTGACAA 16794
280 spGlyLysSerIle...ThrHisGluGlyArgGluIleLeuAlaGluGlu 295
      ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
16793 GGGGTGAACCTGTCAGGCTACTGTGACGACGATGTGACACCTGAAATG 16744
296 LeuCysThrProProAspProGlyAlaAlaPheValValValGluCysPr 312
      :::: ||| ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
16743 GTGTGGAGCCTCAAAAGCCTGGCAAGAGACTGGCAGTAATGACCTCCC 16694
312 oAspGluSerPheIleGlnProIleCysGluAsnAlaThrPheGlnArgT 329
      | ::::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
16693 GACTTCGGATTATGTGAGAGCTGTGTAACCCGCCCGCCGAATGGAAG.... 16648
329 yrGlnGlyLysAlaAspAlaProValAlaLeuValValHisMeLAlaPro 345
16648 ..... 16648
346 AlaSerValLeuValAspSerArgTyrGlnTrpMetGluArgPheGl 362
      ||| ||| ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
16647 ...TCCCCCTCAGTTGCAAGCAATTTGGCAGCTTCTTTTGGATCCTGGG 16601
362 yProAspThr....GlnHisIleValLeuAsnGluAsnCysAlaSerValH 378
      |||| :::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::

```

```

16600 TCTGTGTGTCGGGGAACACCCCTCGGCTTCGGGAATTCGTGCTTCATATGC 16551
378 isAsnLeuArgSerHisLysIleGlnThr..... 387
      ||| :::: ||| :::::
16550 CGAACTGCAAA...CACATCGTCTCGAGCAGCGGATTATTGCCCGCACTAT 16504
388 .....GlnLeuAsnLeuIleHi 393
      ::::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
16503 CTGACCATGCAGAGTGTGTCGGGTGCGGCATTCGCATGCCGCGGATAAG 16454
393 sProAsp.....IlePheProLeu..... 399
      ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
16453 GCTGATAATTAACCTAGTCTCCTATACAGCAACAACAGAGTTTCCACAAA 16404
400 .....LeuThrSerPheArg 404
      ::::
16403 CGCAATCCAGTCTGGAAGATCGCATGCGCGTGGATAGCCAAACCATTTCA 16354
405 CysLysLysGluGlyProThrLeuSerValProMetValGlnGlyGluCy 421
      ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
16353 TTCGAACGAGCCGAGCCGCGGACTCATGCTTAGTAGTATG...GAGCCCGAGTT 16307
421 sLeuLeuLysTyrGlnLeuArgProArgArgGluTrpGlnArgAspAlaI 438
      ::::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
16306 CAAGATAAACTCAGATGAAGTACCGGACGA...CTGAACGCGAGGTAGCA 16260
438 leIleThrCysAsnProGlu.....GluPheIleVal 448
      ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
16259 TCCTGACAAAGATCCCGCAGCTCTGCTAAGGGCGGCTCGACTAGTTACC 16210
449 GluAlaLeuGlnLeuProAsnPheGlnGlnSerValGlnGluTyrArgAr 465
      :::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
16209 AAGAGGCTGCAACATCCGCTACGCCAGGAGAAGATTGAGCAATTTGTGAG 16160
465 gSerAlaGlnAspGlyProAlaProAlaGluLysArgSerGlnTyrProG 482
      ::::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
16159 CGAATTCCTGGTCCCGATGCG.....G 16137
482 luIleIlePheLeuGlyThrGlySerAlaIleProMetLysIleArgAsn 498
      ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
16136 AGATCATCAGCTCGGAACGGGATCTTCGACCTTCCAAATATCGGAAT 16087
499 ValSerAlaThrLeuValAsnIleSerProAspThrSerLeuLeuLeuAs 515
      ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
16086 GTTTCAGCAGCCCTGTTTCATGTTCCAGTGGGGATATTATCTATTGTA 16037
515 pCysGlyGluGlyThrPheGlyGlnLeuCysArgHisTyrGly...AspG 531
      ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
16036 TTGTGGGAGGCGCGCTTGTGTCAGCTTAAGCGCATGTTGAGCCCTGAGC 15987
531 InValAspArgValLeuGlyThrLeuAlaAlaValPheValSerHisLeu 547
      ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
15986 AGCTGCGTGAAGTTCTCAAAACCTCGGCTAATCTCGATTAGCCATCTC 15937
548 HisAlaAspHisThrGly..... 554
      ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
15936 CATCGGAGCCACCCACCTTGGTACAGTCTCTGTTATAAAGGCTTGGTACCA 15887
555 .....LeuProSerIleLeuLeuGlnArgGluArgAlaLeuAlaSerL 569
      ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
15886 GGAAGAACTATCCCGGAGGTGTCACAGCTGAGCGAGCCCTGAAAAGACA 15837
569 euGlyLysProLeuHisPro.....LeuLeuValValAlaProAsnGln 583
      ::::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
15836 TAGGGAAGATCCTGGAGGAAGCGTTTGTGTCGTATCTGATCTCATG 15787
584 LeuLysAlaTrpLeuGlnGlnTyr..... 591
      :::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
15786 ATGATTGAGTGTGGAGGAATATGCTGGTGTGTAAGATTTTGGTTTCGC 15737
591 ..... 591
15736 CAAGCTCACCCCTCTATCTGCCTATCTATACAGACTGGCGACGAGGGA 15687

```

592HisAnGlnCysGln..... 596
15686 TGAATGTCATCAAAATACCCCACTGCCGTGCCGACGGCTCTTATCOT 15637
597 ...GluValLeuHisHisHisSerMetIleProAlaLysCysLeuGlnG 612
15636 GGGCGGGAAGTTGACACATATAA.....CCGA 15608
612 uGlyAlaGluLe.....SerSerProAlaValGluArgL 624
15607 ACAAAGTGAACCTGAATTTAACTGCAAGGATCACAGGA...GAGGAGC 15561
624 euileSerSerLeuLeuLeuArgThrCysAspLeuGluGluPheGlnThrCys 640
15560 TAAGCGCAAGCTTCGCAAGCCAGCGGTCTGTCGGATATTCTAACGGCA 15511
641 LeuValArgHisCysLysHisAlaPheGlyCysAlaLeuValHisThrSe 657
15510 TAGCTCTCCATTGCGGTGGTGAATGGAGTCTCACTCATCTTCCCTGA 15461
657 rGlyTrpLysValValTyrSerGlyAspThrMetProCysGluAlaLeuV 674
15460 TGGCTTCAAAGTCTCTCTCAGGTGACTGCCGGCCATCGCTAGCTTG 15411
674 alArgMetGlyLysAspAlaThrLeuLeuIleHisGluAlaThrLeuGlu 690
15410 TGACTATTGGACAGACACTACCGTCTTATCCAGCAAGCCACATTTAGC 15361
691 AspGlyLeuGluGluAlaValGluLysThrHisSerThrThrSerG 707
15360 GATGACATGGTGGCTCGCTCTCCCAAAAACATTCACCGCCCAAG 15311
707 nAlaIleSerValGlyMetArgMetAsnAlaGluPheIleMetLeuAsnH 724
15310 AGCTATTGAAGTAGCCGCAAAATCGCGCTCGCACTATCTCTCTCACCC 15261
724 isPheSerGlnArgTyrAlaLysValProLeuPhe..... 735
15260 ACTTCAGCAGCGCTACCAAAAATCGCCCACTTCAATCAGCCCAAGAA 15211
735 735
15210 CTCGTCCAGGCAAGTCACTAGTCACAGCCCGTCACTTCAGGGGCACAAAAG 15161
735 735
15160 TATCAGCGCAAGCAAGAAACAGCAGCAGCCGGAAGCAGCATCCACAGATA 15111
735 735
15110 TTCCCTTCAAGAACGACATTTGATGGCGCAGAGAACATTTGCCCTTGATCAA 15061
736SerProAsnPheSerGluLysValGlyVal..... 745
15060 CTGCGGCTTGAACCCCGCAGCCAGCGGAGTGGTTAGAACATATCTGAGCC 15011
745 745
15010 TTCAGAGCTTCCAGAACCGGATGCCCCAGCAGCTCCCATCGTCCCATAA 14961
746AlaPheAspHisMetLysValCysPheGlyAspPheProThrMet 760
14960 TTGCAGCATTTTGACACATGCGCTCGCGCTCCGCGATATGTACACACTC 14911
761 ProLysLeuIleProProLysAlaLeuPhe..... 771
14910 GAACAATATGCACTGCGGTGAGCGCTCTTCGATATCATCATCAGCGCGC 14861
772AlaGlyAspIleG 776
14860 CTCCAAACTCGAACAGAACCAAGCGGAGAAAGCGAGCGGTGAGGTGCG 14811

776 luGluMetGluGluArgArgGlu.....LysArgGluLeu 787
14810 AGGAAACGAGAAAGAGGAGTTTAGCGGGAGAACAAAGTTTAAACA 14761
788 ArgGlnValArgAlaLeuLeuSerArgGluLeuAlaGlyGlyLeuCl 804
14760 AAGCAGGACGATGAGCCAGGAGCAAGGAATGGCAAGCGCGAGAA 14711
804 uAspGlyGluProGlnGlnLysArgAlaHisThrGluGluProGlnAla 820
14710 GGAAGCCCGGTGCTGATAGTCTCTCTCAGTCTCGGCTCAATCT 14662

seq_name: gb_in2:CELE04A4

seq_documentation_block:

LOCUS CELE04A4 22846 bp DNA INV 19-DEC-1997
DEFINITION Caenorhabditis elegans cosmid E04A4.
ACCESSION AF038611
VERSION AF038611.1 GI:2702410
KEYWORDS
SOURCE
ORGANISM

Caenorhabditis elegans strain=Bristol N2.

Eukaryota: Metazoa; Nematoda: Chromadorea; Rhabditida;
Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 22846)

REFERENCE

AUTHORS

Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M.,
Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J.,
Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A.,
Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M.,
Johnston, L., Jones, M., Kerhaw, J., Kirsten, J., Laister, N.,
Latreille, P., Lightning, J., Lloyd, C., McMurray, A., Mortimore, B.,
O'Callaghan, M., Parsons, J., Percy, C., Rifken, L., Ropra, A.,
Saunders, D., Showkneen, R., Smalton, N., Smith, A., Sonhammer, E.,
Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M.,
Vaughan, K., Waterston, R., Watson, A., Weinstock, L.,
Wilkinson-Sproat, J. and Wohldman, P.

TITLE

2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

JOURNALS

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Submitted by:

Genome Sequencing Center

Department of Genetics, Washington University,

St. Louis, MO 63110, USA, and

Sanger Centre, Hinxton Hall

Cambridge CB10 1RQ, England

e-mail: rw@nematoe.wustl.edu and jesus@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone.

It may be shorter because we only sequence overlapping sections

once, or longer because we provide a small overlap between

neighboring submissions.

This sequence was finished as follows unless otherwise noted:

all regions were double stranded or sequenced with an alternate

chemistry; an attempt was made to resolve all sequencing problems,

such as compressions and repeats; all regions were covered by

sequence from more than one subclone

NEIGHBORING COSMID INFORMATION:

The 5' cosmid is K08F11, 200 bp overlap;3' cosmid is T16C6. Actual start of this cosmid is at base position 197 of CELE04A4.

NOTES:

Coding sequences below are predicted from computer analysis, using the program Genefinder (P. Green and L. Hillier, ms in preparation).

FEATURES
source

Location/Qualifiers

1. .22846

/organism="Caenorhabditis elegans"

/strain="Bristol N2"

/db_xref="taxon:6239"

/chromosome="IV"

/clone="E04A4"

3901. .6015

/gene="E04A4.3"

join(3901. .3962,4015. .4289,4457. .4560,4620. .4746,

4793. .5072,5424. .5610,5658. .5768,5815. .5921,5967. .6015)

/gene="E04A4.3"

/codon_start=1

/evidence=not_experimental

/protein_id="AAB92038.1"

/db_xref="GI:2702414"

/translation="MDTAKWFTCFEWEQVIREKKYLEATQVTIRELEMLKEDKLLTL
RVPACNDKLPDYTCRSKQKMGDHHVVRHEPVYKQLFQRTTHVKCTQFETGWIT
RHVELLDKLAITYTAVLKTDVPHRTIGALKVARDNDIPITSKOMQNFGRHPTAVP
KGSPRPATTVHELEKLEKLEDEVMFSIDNSNLAETIYKHFFDAKMLTAGPTEET
ELDWRDEVERICMDSTSRKALKILLERTDGRKARIFLDSFTLDTGYTVVVM
GELNSFTTRVPLAYMIHTDTHAFANWLTDEKNAIGWGRRCPCLLV
DGEAALVYAEKLTDLVRCDVHIMSLMQHYKGVKVLDSKLLFEGECRNTWARGLL
GSFEADPDEKLVETKSPNSDVKVRNFMILILNASVL"

8026. .9173

/gene="E04A4.2"

join(8026. .8046,8464. .9058,9115. .9173)

/gene="E04A4.2"

/codon_start=1

/evidence=not_experimental

/protein_id="AAB92037.1"

/db_xref="GI:2702413"

/translation="MLLKIEIRGLSLDDPLTFEMFDGNGFKFTTRIFONLKGSYVO
KGLTSLQSDSVLIANPDPIOMNAVLCNLHETSLKWFRAQPEDLLISNRYL
SISEDMSKKVNLKELKHWINGSNRMEDVCFVRVFNDDVLKFLKNDLICAPVE
RELIVQPGKEQIVKGYDVRMDGTATVITVDHIOSLKMFVRVNEYIITIIII
YVLEEA"

10829. .12409

/gene="E04A4.1"

join(10829. .10920,11090. .11970,12153. .12409)

/gene="E04A4.1"

/codon_start=1

/evidence=not_experimental

/protein_id="AAB92036.1"

/db_xref="GI:2702412"

/translation="MVDVSVFPOLLKLPKSSNVVHRMLVADLIGFSLTSKNKLO
TKRLGLKAEDVHLEIRSGIILHISLOHTFTFYKTYNEYRSEHREELNMYLCTR
LGNKFTIYVRQKEDVLDVDFHVNINISELEITTDHFTIYDLEIESFSWYHI
DLNKLQYARLKRFSKRANQSLILMSRNPTAETSQLOVVYTRNMDHIVLYHPKIYD
LNDLVNTVSLYMDKIEDVMELKHWIGDLKLELACFGYKELQKNAAVVEILF
KNVDYQVAPARELTIIPTGSOQICVKGVDVRRVDGTEENITVNSORKCLTKIPGSL
FESVHFYCVCATNQRNLYLLEFIKIMKIYTSNAEKSFDAHNTETPKNSIKHERRD
PRPIENVPPNLLIGG"

complement(12921. .17200)

/gene="E04A4.4"

complement(join(12921. .13023,13237. .13460,13999. .15037,
15263. .15402,15449. .16319,16572. .16703,17175. .17200))

/gene="E04A4.4"

/note="contains similarity to sulphatases; coded for by C.
elegans cDNA yk19h5.3; coded for by C. elegans cDNA
yk39c9.3; coded for by C. elegans cDNA yk15d11.3; coded
for by C. elegans cDNA yk42h12.3; coded for by C. elegans
cDNA yk22g6.3; coded for by C. elegans cDNA yk19h5.5;
coded for by C. elegans cDNA yk22g6.5; coded for by C.
elegans cDNA yk15d11.5; coded for by C. elegans cDNA
yk42h12.5; coded for by C. elegans cDNA yk39c9.5"

alignment_scores:

Quality: 591.50

Ratio: 1.259

Length: 1038

Gaps: 35

/codon_start=1
/protein_id="AAB92042.1"
/db_xref="GI:2702418"
/translation="MKMLFFGKIKVSRHLISSTCLFKNNEELLESIKERIARNRRL
QKHSSEHLKAREVNASISLRQSMAAVOKKOKAAHEPPANSIVNIPSOVSIEVLNGT
GLLRACFILRPLTKTYMFCNACRFLWOLIRSSSVVDLFTISANWNTIACISIL
LSKSEPLNLHGMANKHLECIKRFQDSDYGSCKYPSQVEERYPTMENTEDAGLK
VYIPLSPNLIGSNNEKSNKVNNDVIAFLIENKEAARIDTMMKLMELKVPKGPLI
GKLKSGEATVLPDGRITQIDQVFSDDKGVKPLLLVTECTEDHVKALIDSSSLQPF
LNGEKOLDVHVISDDAVINTPTYRHLMEKLNPSITILLINGNPVTPAVESVYKHT
RLLSRIASPLPALPIDWSGIITONEELSORQDQIRVAPMORVMRERGASNEEPI
VNNLLAEPESLSDKAKELIKYQKLEKENKMDCEPKLFTFCSTSAVPSKYRNVTYL
VEANSAIILDVGGTYGQMRVFEQCKOLLVNLKCLITHAHQDHMNLGTYIIA
RRKEFESLGAPRPLVLCNRRNLKPMKTYISCFENIEHLEIIVDISRYLTPPCSP
GGPKGRPLPSPLPPSRDVLQDMSSFDKAKMLDELKAVQVHITRMANGFVRVA
GKRIVSGDTKPCDLLVEGKADVLVHESFEDGHEVDMPKPKLAKISLADAM
RKRIVSGQAVDVGKRMNAKHLILTHFSARVPKVPVLPVLDKENGKGMVMDLRVFD
HLPLVSKLLPIPREVFVAELFELTIKKEQRLVKDLSEKRLQKLA"
complement(17983. .18738)
/gene="E04A4.5"
complement(join(17983. .18093,18151. .18261,18318. .18503,
18553. .18664,18713. .18738))
/gene="E04A4.5"
/note="contains similarity to S. cerevisiae mitochondrial
import protein (MIM17) (GB:X87371); coded for by C.
elegans cDNA yk83e7.5; coded for by C. elegans cDNA
yk211c5.5; coded for by C. elegans cDNA yk32a7.5; coded
for by C. elegans cDNA yk465g9.5; coded for by C. elegans
cDNA yk81g5.5; coded for by C. elegans cDNA yk25e2.5;
coded for by C. elegans cDNA yk11e10.5; coded for by C.
elegans cDNA yk400c6.5; coded for by C. elegans cDNA
yk452a5.5; coded for by C. elegans cDNA yk452a5.3; coded
for by C. elegans cDNA yk211c5.3; coded for by C. elegans
cDNA yk502c12.3; coded for by C. elegans cDNA yk465g9.3;
coded for by C. elegans cDNA yk32a7.3; coded for by C.
elegans cDNA yk11e10.3; coded for by C. elegans cDNA
yk83c7.3; coded for by C. elegans cDNA yk25e2.3"
/codon_start=1
/protein_id="AAB92039.1"
/db_xref="GI:2702415"
/translation="MEETREPCPYRIGDDIGSAPAMGLVGSIFQAFGGYKNAAGK
KLVGMRVRRMRSTLTGVOFAFGMFTDCCLVAIRKKEPDIINSIVSGGLTGALLA
IRSGPKVWAGSAILGSVILAMTEGQVLVTRMMGAMMDPTOPPEALDDPSRLGQSK
AEPGLDPRFPFGIPTGLPNLS"
complement(18757. .18758)
/note="5' splice site; see EST yk83e7.5"
complement(19059. .21097)
/gene="E04A4.6"
complement(join(19059. .19084,19138. .19318,19365. .19496,
19547. .19794,19849. .20377,20760. .20863,20917. .21097))
/gene="E04A4.6"
/codon_start=1
/evidence=not_experimental
/protein_id="AAB92040.1"
/db_xref="GI:2702416"
/translation="MGTAFFNRNLFVNHLLGFLITLYCSADTPVEYCPVSWRHRI
DSIFPTSLFCVKNDDLEVDIVGEOWNEFESKRWQDSDSHSENLELLHFG
SSVRLKFKFPAAISASDECTGCTQNTVITKRNMTVKCKPLATDILSDSENLELLHFG
LFGNGDFVERTENETIAPRISTAQFIAPVWNSMCYGLKANSIRFNGIKTNSVE
IDAQVDSLPGNVNDNNWFQTFVSVMWIPVASEIETNTVGVKAGEQYRVKGSIPV
PFAVPTLGCYSGSIAPSVFRLSRMSVCTISTINCEADAKARAFYBQVYPSVLVS
SPSEDAHPASVGRVNTWELSPSTISCRLLPVSSLLQIYVSKQSTKNRYVITAGNS
QLLLDDVPYLSGQETRMPISISFTEVTPPKNIFSPALPYVDIRLPHDFPFPMTSTNY
SASNYSQLFILVNEY"
complement(21320. .21706)
/gene="E04A4.7"
complement(join(21320. .21475,21527. .21706))
/gene="E04A4.7"
/note="C. elegans cytochrome c (SW:P19974); coded for by

Percent Similarity: 45.279 Percent Identity: 21.484

alignment_block:

US-09-434-382-2 x CELE04A4/rev ..

Align seg 1/1 to reverse of: CELE04A4 from: 1 to: 22846

```
55 ProAsnThrValTyrLeuGlnValValAlaAlaGlySerArgAspSerGI 71
||||| ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
16210 CCCAGTCAGTATCGATCGAAGTCTCGGTAATGGAAACAGGACTTCCTCG 16161
16160 AGCATGCTTCATCCTTCGAAACACCACTGAAACCTACATGTTCAACTGCC 16111
71 yAlaAlaLeuTyrValPheSerGluPheAsnArgTyrLeuPheAsnCysG 88
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
16160 AGCATGCTTCATCCTTCGAAACACCACTGAAACCTACATGTTCAACTGCC 16111
88 lyGluGlyValGlnArgLeuMetGlnGluHisLysLeuLysValAlaArg 104
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
16110 CTGAGAATGCATGCAGATTCCTGTGCAGCTTAGGATCCGAAAGTTCACGC 16061
105 LeuAspAsnIlePheLeuThrArgMetHisTyrPheSerAsnValGlyClyLe 121
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
16060 GTCTGTGATCTTTTCATCATCACTGCAAAATGGGATAATATTGCTGGAAT 16011
121 uSerGlyMetIleLeuThrLeuLysGluThrGlyLeuProLysCysValI 138
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
16010 TTCATCAATTCATCTTCT...AAGAAATCAAAATGCATGCTACTCGAC 15964
138 euSerGlyProGlnLeuGluLysTyrLeuGluAlaIleLysIlePhe 154
||| ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
15963 TTCATGGTGCAATGAATATCAAAACACTTTTGGAAATGATTCGACCATTT 15914
155 SerGlyProLeuLysGly.....IleGluLeuAlaValArg 166
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
15913 CAAGATTCGGATTACGGTAGCTGCAAGATATCCATCTCAAGTTGAAGAAG 15864
166 gProHisSerAlaProGluTyrGluAspGluPheMetThrValTyrGlnI 183
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
15863 TCATATCAATGAGAGATATGAGGATCCCGGATTAAAGGTCACATATA 15814
183 leProIleHisSerGluGlnArgArgGlyLysHisGlnProTrpGlnSer 199
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
15813 TTCCCTCG.....AGT 15803
200 ProGluArgProLeuSerArgLeuSerProGluArgSerSerAspSerGI 216
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:
15802 CCA.....CCGTTGAAT.....ATCGGAAGTAAACAAGA 15774
216 uSerAsnGluAsnGluProHisLeuProHisGlyValSerGlnArgArg 233
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
15773 AAAAAGTAAAAAT..... 15761
233 lyValArgAspSerSerLeuValValAlaPheIleCysLysLeuHisLeu 249
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
15760 ..GTAAAAGTCAACAATAGACATATGCAATTTTAAATCGAAATGAAGGAG 15713
250 LysArgGlyAsnPheLeuValLeuLysAlaLysGluMetGlyLeuProVa 266
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
15712 GCAGCCAGCAAGATCGATACCAAGTGAAGTTGATGGAACCTGAAGTACC 15663
266 lcGlyThrAlaAlaIleAlaProIleIleAlaAlaValLysAspGlyLys 283
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
15662 AGGT.....CCACTAATTTGGAACCTGAAATCAGGAGAG 15628
283 erIleThr...HisGluGlyArgGluIleLeuAlaGluLeuCysThr 298
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
15627 CAGTTACTCTTCCAGATGCAAGGAGGATTCACACCTGATCAAGTTTCTCA 15578
299 Pro.....ProAspProGlyAlaAlaPheValValValGluCysPr 312
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
15577 TCGGATAAAGTTGAGGAGACAAGCCACTGCTTTTAGTCAGACATGTAC 15528
312 oAspGluSerPheIleGlnProIleCysGluAsnAlaThrPheGlnArgT 329
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```
15527 TACAGAGGATCATGTGAAAGCACTGATAGACTCATCATCTCACTGCAACCAT 15478
329 yrGlnGlyLysAlaAspAlaProValAla..... 338
||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
15477 TTTTGAATGGGAAAAAACAGTTGGATTATGTAAAGATTGAAAAAGAATTA 15428
339 .....LeuValValHisMetAlaProAlaSerV 348
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
15427 TTGACATTTAAAAAATACATTTTCAGATGGTTTCATATCAGTGATGTCAG 15378
348 alLeuValAspSerArgTyrGlnGlnTrpMetGluArgPhe...GlyPro 363
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
15377 TATCAATACTCAACCTACAGACTTTGATGGAGAAGCTCAACAATCCC 15328
364 AspThrGlnHisLeuValLeuAsnGluAsn.....Cy 374
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:
15327 AGTATCACTCATTTGCTGATCAATGGAGGAATCCTGTGATTCACACTGT 15278
374 salaSerValHis.....A 379
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
15277 GGAAGTGTCTATAAGTAAGGAAATATGAAATATTGAATAGATCTACGA 15228
379 snLeuArgSerHisLysIleGlnThrGlnLeuAsnLeu..... 391
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:
15227 ATTCGAGATCTCGAAACAAGATCCAGGGTCTCAAGAGCTGCAGTTGC 15178
392 .....IleHisProAspIlePheProLeuLeuThrSer... 402
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:
15177 ACGATTTCCAGGACCTCTGTTTTCGAGATCTTCTATTGACGAGATCCCG 15128
403 .PheArgCysLysLysGluGlyProThrLeu...SerValProMetValG 418
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
15127 TTTTCGTGTTCCCTCGCGGCATCAGAAAAATTCAAAAAATTCAGTTTAC 15078
418 lngGlyGluCysLeuLeuLysTyrGlnLeu.....ArgProArgArg 431
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:
15077 AAATCCATAGTTTAAAGATCCAAATTTATTATTACAGACACACTCGC 15028
432 .....G1 432
15027 CTCTTCGATCAATTCCTCATCTCTTCCAGCTCTTCCCAATTCGA 14978
432 utrP.....GlnArgAspA 437
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:
14977 TTGGAGTGGAAATTATACTCAAATGAGGAGCTATCAACACGACGAGATC 14928
437 laIleIleThrCysAsnPro..... 443
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:
14927 AATTCATTCGAGTGCCTCAATGCAACGATATTGGATGCGCGGAGAGCT 14878
444 .....GluGluPheIleValGluAlaLeuGlnLeuProAsnPheG1 457
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
14877 TCGTTCAACGAGGAGCCCATCGTCAACAATTTGCTAGCTGCTGACCCAGA 14828
457 nClnSer.....ValGlnGluTyrArgArgSerAlaG 468
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:
14827 GCTGTCGGATAAAGCGAAGGAATTGATTAAAGGAATATCAGAAGTTGGAGA 14778
468 lnAspGlyProAlaProAlaGluLysArgSerGlnTyrProGluIleIle 484
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
14777 AGGAGATAAAATGCGACTGCCGAA.....TTTCCGAAATTAACA 14740
485 PheLeuGlyThrGlySerAlaIleProMetLysIleArgAsnValSerAl 501
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
14739 TTCCTCGAATCTGTCGCGGTTCCTCAAAATATCGAAATGTGACAGG 14690
501 aThrLeuValAsnIleSerProAspThrSerLeuLeuLeuAspCysGlyG 518
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
14689 ATATCTGTGGGAAGCATCAGAGACAGTCCGATTTTAATTTGATGTTGGAG 14640
518 luGlyThrPheGlyGlnLeuCysArgHisTyrGly...AspGlnValAsp 533
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
14639 AAGGAACATATGGCAAAATGAGACAGCTTTTGGAGAAGATGCTGTTGAAG 14590
```


[illegible]

```

25316 AGGAGAAATAAATGGACTCGAA.....TTTCGGAATTAACA 25279
485 PheLeuGlyThrGlySerAlaIleProMetLysIleArgAsnValSerAl 501
25278 TTCTTCGGAACCTCGTCCGGTTCATCAAAATATCGAATGTGACAGG 25229
501 aThrLeuValAsnIleSerProAspThrSerLeuLeuLeuAspCysGlyG 518
25228 ATATCTGGTGGAGCATCAGAACAGTCCGATTTAAATTGATGTTGGAG 25179
518 luGlyThrPheGlyGlnLeuCysArgHisTyrGly...AspGlnValAsp 533
25178 AAGGAACATATGACAAATGAGAGCAGTTTTCGAGAAGATGTTGTAAG 25129
534 ArgValLeuGlyThrLeuAlaAlaValPheValSerHisLeuHisAlaAs 550
25128 CAGCTTCTGTAATCACTCACTCGTTCCTATCACTCAGCCTCATCAGGA 25079
550 pHisHisThrGlyLeuProSerIleLeuLeuGlnArgGluArgAlaLeuA 567
25078 TCATATGAATGGCTGTACACAATTTATTCACGGAGAAAGAGCTTTTCG 25029
567 leaSerLeuGlyLysProLeuHisProLeuValValAlaProAsnGln 583
25028 AGTCCCTCGAGCCCTTACCGTCCACTAGTCTCTGTATGCAATCGGAAT 24979
584 LeuLysAlaTrpLeuGlnGlnTyrHisAsnGlnCysGlnGluValLeuHi 600
24978 GTTCTGAAACCAATGAAGACCTAC...TCAATTTGCTTTGAAACCATTTGA 24932
600 sHisIleSerMetIle..... 605
24931 GCATCTCTCGAAATTTGTTGATATCTCCAGGTATCCACTAACACCTCCCG 24882
606 .....ProAlaLysCysLeuGlnGluGlyAlaGluIle 616
24881 GGAGCCCTGGGGCCCTCTCGAAAG.....ACCGAGACTTCGGAGGCC 24836
617 SerSerProAlaValGluArg...LeuIleSerSerLeuLeuArgThrCysA 633
24835 ACATCTCCCGCCGAGCAGAGATGTTCTACAGGATATGTCAGCAGTTTCG 24786
633 sp.....LeuGluGluPheGlnThrCysLeuValArgHis 644
24785 ATAAAGGCATGGAAGCTGGATGAGCTGAAGGCTGTACAAGTTTCATCAT 24736
645 CysLysHisAlaPheGlyCysAlaLeuValHisThrSerGlyTrpLysVa 661
24735 ACNAGATGGCAATGGA...TTTGTGATGAGAGTTGCTGGAAACGAAAT 24689
661 lValTyrSerGlyAspThrMetProCysGluAlaLeuValArgMetGlyL 678
24688 TGTATTCTCTGGAGATACGAAACCATGTGATTTGTTGGTTGAAGAAGAA 24639
678 yAspAlaThrLeuLeuIleHisGluAlaThrLeuGluAspGly..... 692
24638 AAGATGCTGATGTTCTGTTTCATGATCACTCACTTTTGAAGATGTCATGAG 24589
692 ..... 692
24588 GTAGATATGACACCAACCCCAAAAAAATTTGGCTAAAAATTTTCGTCTTT 24539
692 ..... 692
24538 GGTTTATTACTAAATGTACGAGATTTTATAGCCAAATACGGTAGCTAG 24489
693 .....LeuGluGluGluAla 697
24488 GGTCTCAACACGACAAATTTTGGATAAATGACGAGATTTGAAGATAACAGA 24439
698 Val..... 698
24438 ATTTTCATCGAATTTTCATAGTTTTTCGTTGAGAAATATATATAAATCTGTG 24389

```

```

698 ..... 698
24388 AGAAAAATCGATGAAATCTTGAGAAACAAAAAGTTTGAACCTACAGTAT 24339
698 ..... 698
24338 TCTTCAAGCGGAATAACATTTTTCGATTCGGCAAAAATTTTTCGTGTCG 24289
698 ..... 698
24288 AGACCCCTATATTACGTATTTTAGTGCAATAATGAAAAATCCAACATTATTA 24239
698 ..... 698
24238 GAAACCCGCTTTTTCAGTTTTTTTTTAAATTTGAATTCATCCAAACTG 24189
698 ..... 698
24188 AGAATTCAGACCAATTCAGGACTTGCTCTCTCATTCGCGCTGATTTGG 24139
698 ..... 698
24138 TCGAATATGCGCGGAGCTATTTCGCTGATTGTCATATTCTCATTTGGC 24089
698 ..... 698
24088 AGCATTTCAAACAGCAATTTTCTGAGTTTCGGACTATTTTGGTTTTT 24039
698 ..... 698
24038 TAAGCAAAATATTGAAATTTTCAACCAATTAATTTTCAGGCTGATGCAAT 23989
699 GluLysThr.HisserThrThrSerGlnAlaIleSerValGlyMetArgM 715
23988 GAGAAACGTCATTCACAAATGGGTCAAGCAGTCGATGTTGGAAACGAA 23939
715 etAsnAlaGluPheIleMetLeuAsnHisPheSerGlnArgTyrAlaLys 731
23938 TGAATGCCAAGCACATATTCTGACACATTTTCAGTCGAGATATCCAAAA 23889
732 ValProLeuPheSerProAsnPhe.....SerGluLysValGlyValAl 746
23888 GTACCACTTCTT...CCTGAATATCTGTATGAAGAAATATTGTTGGC 23842
746 aPheAspHisMetLysValCysPheGlyAspPheProThrMetProLysL 763
23841 GATGGATATGTTGAGAGTTCGATTCGATTCATTTACCACCTGTTTCGAAC 23792
763 eullePro 765
23791 TTCTTCCC 23784
seq_name: gb_pl2:AP001550
seq_documentation_block:
LOCUS AP001550 143209 bp DNA PLN 29-MAR-2000
DEFINITION Oryza sativa genomic DNA, chromosome 1, PAC clone:P0431F01.
ACCESSION AP001550
VERSION AP001550.1 GI:7340902
KEYWORDS
SOURCE
ORGANISM
Oryza sativa (cultivar:Nipponbare) DNA, clone:P0431F01.
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.
REFERENCE
1 (bases 1 to 143209)
AUTHORS Sasaki,T., Matsumoto,T. and Yamamoto,K.
TITLE Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
clone:P0431F01
JOURNAL Published Only in Database (2000) In press
REFERENCE
2 (bases 1 to 143209)
AUTHORS Sasaki,T., Matsumoto,T. and Yamamoto,K.
TITLE Direct Submission

```

JOURNAL

Submitted (27-MAR-2000) to the DDBJ/EMBL/GenBank databases. Takuji Sasaki, National Institute of Agrobiological Resources, Rice Genome Research Program; Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan (E-mail:tsasaki@abr.affrc.go.jp, URL:http://www.dna.affrc.go.jp:82/, Tel:81-298-38-7441, Fax:81-298-38-7468)

COMMENT

The orientation of the sequence is from SP6 to T7 of the PAC clone. Genes were predicted from the integrated results of the following:GENSCAN1.0, BLASTN2.0, BLASTX2.0 as well as SplicePredictor (October1998 version). The genomic sequence was searched against the non-redundant database NRP(PIR,SWISSPROT, GENPEPT, PDB) from MAF, DNABank and the cDNA sequence database at RGP. Protein similarities of the coding regions were searched against NRP with BLASTN2.0. ESTs represent the identified cDNA sequences using BLASTN2.0 with the corresponding DDBJ accession no. and RGP clone ID. This sequence of P0431F01 clone has an overlap with P0485D09 clone at the 5' end.

The sequence of this clone starts at the position 93,219 of P0485D09.

Detailed information on assemble quality together with annotation of this entry at <http://www.dna.affrc.go.jp:82/genomicdata/GenomeFiniished.html>.

FEATURES

source

Location/Qualifiers

1..143209
 /organism="Oryza sativa"
 /cultivar="Nipponbare"
 /db_xref="taxon:4530"
 /chromosome="1"
 /clone="P0431F01"
 /complement(join(10340..14224,14755..14802))
 /note="Similar to AF049110.1(AF049110|pid:94206306) Zea mays retrotransposon Cinfu1-1; internal stop codons:13700..13702,13391..13393,13272..13274, 13086..13088,13020..13022,11549..11551,11558..11560, 10880..10882; probably inactive because stop codons and frameshift positions are included in CDS"
 /pseudo
 /codon_start=1
 10917..10918
 /note="one bp frameshift deletion"
 11429..11430
 /note="one bp frameshift deletion"
 12388..12389
 /note="one bp frameshift deletion"
 12743..12744
 /note="one bp frameshift deletion"
 13290..13291
 /note="two bp frameshift insertion"
 /complement(join(17293..17493,18173..18283,18920..18973))
 /note="hypothetical protein"
 /codon_start=1
 /protein_id="BAA92975.1"
 /db_xref="GI:7340903"
 /translation="MPTVQIIPRTGMVFILNFDLIASSVCVCANMOCIMAGERST DGLFFPAGQDAEKQKQAKDDLSRGLCPVSCMPLHADNDVVGSDFWA AGGGGGGAPPLAGNNLR"

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

CDS

join(33172..33273,33899..34024)
 /note="hypothetical protein"
 /codon_start=1
 /protein_id="BAA92977.1"
 /db_xref="GI:7340905"
 /translation="MTIFRYCYCEKVVMLHDDCYVTAKFLRDVNNQVISISTIGDGR PFNTCKIRLIMFISRPVHRSLAGPHEQ"

CDS

complement(join(34876..35013,36635..36695,36723..36861, 37054..37169,38057..38157))
 /note="hypothetical protein"
 /codon_start=1
 /protein_id="BAA92978.1"
 /db_xref="GI:7340906"
 /translation="MHTIRCISAGALVLMVNLNKEELYVVGSSNOFFNSOHLSP HLYLLSSPHTPPSLSNLPVRAAAGGHTVAVLHPAAGGSRSTAEVARGVWRK EEAANAIPVDLAADGFFSRADDISTRTKTKTDFSVMDGISSTGSEQIELAAPSS APASELGSHPSSRRRRRHDPGTPA"

CDS

join(40151..40508,41020..41387)
 /note="hypothetical protein"
 /codon_start=1
 /protein_id="BAA92979.1"
 /db_xref="GI:7340907"
 /translation="MGTEKKHQQRSRRRRKDVTKGSRCEVASKASDSDGVQRKH KQGYTRTLPSPFLLYLDRTLOPDASLAFAHFPLCAPTRSSARSELARCCAC FGFGVGQGEYERCHSVPGGLELGDTPPTRLAMVGTSSGNLSNGDNDYGSIFS LKSPLAWTOPPPSLRTGSSMTMRCDGLDLPDPTTDRSTATIGVWRGSRGGGARR GNSLDDDELGLGDGATADER"

CDS

join(42260..42269,42373..42761)
 /note="hypothetical protein"
 /codon_start=1
 /protein_id="BAA92980.1"
 /db_xref="GI:7340908"
 /translation="MCTSGAGKSSFOASDPSARRVLRTYITCRPAGESIDAAPCCTTR DVPRTGYERPSPAMEHTLSLKSAKYKGFYKRFMAVYLKLSLTAVCTVGNVNRV WGVPPAKTDLAKMLQFLAAHFKTPLAKI"
 /complement(join(46089..46586,47173..47257,48631..48762, 49150..49229))
 /note="hypothetical protein"
 /codon_start=1
 /protein_id="BAA92981.1"
 /db_xref="GI:7340909"
 /translation="MNNAKFYNNMKRKYKISPRNDGVQDRSCLIKDLFGGOLLETPP SASIKVNLRLDVTYHNSQGMVQYIRSSNRDDTPRIYQKFGSDRSTSVSNCDQRPD ITHIELSIPDELPPSTRAGVAVAYAWOPSOHFLKDGAAHLSVTLSSLPLALS LSLRRGHSAAATGRRRAGHGGVGGVGRAGGCGGCGGCGVCGDPSNC PPRDGSILPRPPAMPCTTTAALAAPYPYRGALAREKEEQ"

CDS

join(50116..50125,50292..50697,51472..51704,52558..52654, 52775..52836,53717..53754)
 /note="ESTs AU083004(E60493),AU030968(E60493), D23864(R0457) correspond to a region of the predicted gene.: Similar to Glycine max GH1 protein (AF016633)"
 /codon_start=1
 /protein_id="BAA92982.1"
 /db_xref="GI:7340910"
 /translation="MTYGIQTRIPVHVLPMLPSPPLELDYIGLSPPPPPSSSSAA ARADDYDGLKGTLELRLLPGSESPDRRPAATAAAATATLLELPAGKAKVFPDEAL TPPTAAGKGAAREGEVECAEDKVAAPPOPAKAOVVGPPIRSYKNTMATNO IKSNEEDVAKOCQGLYKVSMDGAPYLKKVLDLKYKNYKDKMSLGLKMFIFGSGK EAGENQDGEYVLITYEDKDDMLVGVDPWEMFTDCRRRLRMKSGSDAIGLAPRAGEK SKNRN"

CDS

complement(join(54434..54639,56954..57117,57172..57379, 57474..57483))
 /note="hypothetical protein"
 /codon_start=1
 /protein_id="BAA92983.1"
 /db_xref="GI:7340911"
 /translation="MRDLSLCOEKKKKREREAPAROLGSGTVECVQTVRFA ROVSGSSHVGTAGAHQVQNPSPRRTCTPTPIFALTSGPTVVVSVLGMCTCSWG ADVDRSSVPTVSDAVGPRVVRGALHAQLADTNSNGKEKEKERTENSIGGK QKOQILQARGGKESLILRDHPSLFIHLYFYTT"

CDS

complement(join(60274..60854,60878..60896,61212..61688))
 /note="hypothetical protein"

CDS

complement(join(60274..60854,60878..60896,61212..61688))
 /note="hypothetical protein"

PGARACAEAVQLURASHRCCSSLAARAPAAATCCRAPPTRRRALARA SPAHRRRTPL
 PGACGAEAGDHMRRCFNQNDIISSRLSRPKKRLPSLHVNNVAIPEELGRPMRR
 RGSVDMGRKSCALCFVLCSESPRIELTYTNRTKNRTSMLSGGELKSSSL
 VOQMVVGTGNSSSSIMGSLRQLPCSEPDAASSPASMFLPQQLILLHASNSPCL
 NPEVNLSTGLDDEAMPHEASAAAYDFAVGGAGDGILOASPEASSCKSLSOMLL
 QAAASPRCVTTSGGSSMDPNSNTAAVAPAEPELTRKHAGGSDNSSEVKEKRLG
 DRTALHQIVSPGKALSYPLGQCCSNPMQOOTVKPQNPFIHTNRVPCWSAAAH
 NGARACAEAAHRRGHSLSRCSRWSSLAEEVQPLVVGRRRADFLSPQELVPL
 PGARACAEAVQLURASHRCCSSLAARAPAAATCCRAPPTRRRALARA SPAHRRRTPL

/codon_start=1
 /protein_id="BAA92984.1"
 /db_xref="GI:7340912"
 /translation="MGSSHAGGEMCMGGGGRARRAGGKKAASAAAAAQHKKVAKOP
 QRLGVLAQLKRLRHNMVAALRSAAAGDAPYSPPPQPOPFPFPHLPLOQDCYE
 AADHRRLAAGGVOPVYEGMLPYSGRLAASPFAVAYEVKGDHGHGOYSSSEQOROP
 OYVSMSSSDYDVGSGSSAGSEELDLELRVLLQOFTTASTPHHLLOCGRGRR
 AREATTGRRRTASGGGAEPHAEAGGEPHROAPRGORPRRAAACAAACGARRP
 AGGAPRPAAGVEARRRRAAASVARSAAVTTGSGRRRAEAAAGAAARPAATSR
 GGGRRRAEAGTWLAASAGGRRSP"
 join(63452..63622,63677..63871,63966..64035,64321..64664)
 /note="EST D15617(C0960A) corresponds to a region of the
 predicted gene.; Similar to Arabidopsis thaliana
 chromosome 4, BAC clone F184; vacuolar sorting
 receptor-like protein (AL021637)";
 /codon_start=1
 /protein_id="BAA92985.1"
 /db_xref="GI:7340913"
 /translation="MALVAAAANQKQKASIGRRWRLRLRLAVLWARKGSVHSLCL
 FSNURRAGLVGVTEETSSARVLRILPICAPVPTPGFYGDEDRYFFCRWDTEPCS
 GVGCYDIENDLVKTEQIVQVHGARGDVTILPTLVNNMOYRAGGKHRGAHGANG
 VARGAHRVRSRGELRRPEDAGADADAEEEDADDHREKQGTTHRRCCGGGG
 GHRPPCSPACRRSRAPLLHSASPAQAQHRSPVPVPPR"
 join(65928..66003,66135..66175,66769..66838,66930..67078,
 67573..67668,67756..67826,67940..68001,68081..68144,
 69038..69122,69207..69333,69446..69534)
 /note="ESTs D41826(S4655),C22685(S4655) correspond to a
 region of the predicted gene.; Similar to Arabidopsis
 thaliana chromosome 4, BAC clone F16A16; protein
 kinase-like protein (AL035353)";
 /codon_start=1
 /protein_id="BAA92986.1"
 /db_xref="GI:7340914"
 /translation="MDRIYGGFKLGRKIGSGFEIYVLAHVDYETAAVKIESKKT
 KPOLFYEAQKLYNALGVGQGEENVLIDLLGPSLEDLFVYVCGKRFLL
 KTVLMLADOMITRTEFMSKGLHDKPDLKMGKQKANOVIIDFGLAKRYRDS
 TNRIHPRENKLNLTARYASCNTHLGVEQSRDDLESGLVLLYFLRGLSLPWGLKA

CDS

CDS

alignment_scores:

Quality: 487.50 Length: 588
 Ratio: 1.812 Gaps: 15
 Percent Similarity: 45.748 Percent Identity: 25.680

alignment_block:

us-09-434-382-2 x AP001550

Align seg 1/1 to: AP001550 from: 1 to: 143209

407 LysGluGlyProThrLeuSerValProMetValGlnGlyCysLeuLe 423
 ||||| |||
 130696 AAGAAGTTCCTCGAACAGATTAGCAAGTTCTGCGCAAGTAATTCAGTTGG 130745
 423 uLysTyGlnLeuArgProArgGluTrpGlnAlaGspAlaIleIer 440
 ||||| |||
 130746 CAAGCATATGCTCATGTTGGAAGCCGATGATTACTGAGAATTCCTCTG 130795
 440 hrCysAsnProGluGluPheIleValGlnAlaLeuGlnLeuProAsnPhe 456
 ||||| |||
 130796 TTGTGAAT.....TTGTAGACGCAAAATAGTAATTACAGTAAGTTA 130836
 457 GlnGlnSerVal.....GlnGluTyArgArgSe 466
 ||||| |||
 130837 CAAGATGGCACTCCATTGAGCAAGTGGCTGGGAAGCAAGCCCAAGGA 130886
 466 rAlaGlnAspGlyProAlaProAlaGluLysArgSerGlnTyPro...G 482
 ||||| |||
 130887 CACTCTGATATCCCTGCTGTTGGAGATGCAACAAGAGAAGATATCG 130936
 482 luIleIlePheLeuGlyThrGlySerAlaIleProMetLysIleArgAsn 498
 ||||| |||||
 130937 AAATTACCTTTTGGTACGGGTTTCATCCCAACCTTCMAAATATCGAAC 130986
 499 valSerAlaThrLeuValAsnIleSerProAspThrSerLeuLeuAs 515

||||| |||
 130987 GTAAGTTCATATATATACCTGTTTACCAAGAGGTATACTTCTTGA 131036
 515 pCysGlyGluGlyThrPheGlyGlnLeuCysArg..... 526
 ||||| |||||
 131037 TTGCGGTGAAGAACCTTAGGGCAATTTGAAAGAAGGTTCATTACAT 131086
 526 526
 131087 ACTTCTTTAGGTAATTTAGAAAATGCATGTTATTTCATATATTGTCGTCATG 131136
 527HisTyRGly....AspGlnValAspArgValLeuGlu 537
 ||||| |||||
 131137 ACATTCTCTTTATAGGTTTGGTAAAGTGGTGCATGATGCTGTGAA 131186
 537 yThrLeuAlaValPheValSerHisLeuHisAlaAspHisHisThrG 554
 ||||| |||||
 131187 AGTTTGAAGTGTATTTGGATTTCCTCATATTTCATGCTGATCACCACACAG 131236
 554 lyLeuProSerIleLeuLeuGlnArgGluArgAlaLeuAlaSerLeuGly 570
 ||||| |||||
 131237 GTGTAGCTAGAGTTCTTGTGCTTTGAGGCTAAATTTGTTAAAA.....GGG 131280
 571 LysProLeuHisProLeuLeuValAlaProAsnGlnLeuLysAlaTr 587
 ||||| |||||
 131281 GTACCTCAACAGCCCTTACTTGTATTGGACCAAGACCCTTCAAAAGATT 131330
 587 pLeuGlnGlnTyHisAsn.....GlnC 595
 ||||| |||
 131331 TCTAAATGCATATTCAACACTCGAAGATCTTGATATGCAGTTCTCTAGACT 131380
 595 ysglnGluValLeu..... 599
 ||||| |||||
 131381 GCAGGCAGACTTTAAAGCCTAGTATAGAGGCTTTCTCAGTGACAATGCT 131430
 600HisIleSerMetIleProAlaLysCysLeuGlnGluGlyAl 614
 ||||| |||||
 131431 ACTGAATCTGCCACATCTCAACTTGAAGCACTATATTTGCTCCTCGAAG 131480
 614 aGluIleSerSer.....ProAlaValGluArgLeuIleSerS 627
 ||||| |||||
 131481 TAAATGGAGAACTACAGCGGAGGCTGCAAGTCCAGAGACTACAG 131530
 627 erLeu.....LeuArgThrCysAspLeuGluPhe 637
 ||||| |||||
 131531 CTCTAACTAATTTGAAAGATGCTCATGAATCAGGTTTGGAAAGTTCTT 131580
 638 GlnThrCysLeuValArgHisCysLysHisAlaPheGlyCysAlaLeu... 653
 ||||| |||||
 131581 TACAGTGTCCCAGTTCTGCACCTCCACAGGCGTTTGGAGTTCTTCTGAG 131630
 654ValHisThrSer.....GlyTrpLys 661
 ||||| |||||
 131631 AGCAAGGAGAGGCTGAGAGTCTCGAAAAGCTATTCAGGATGGAAGG 131680
 661 alValTySerGlyAspThrMetProCysGluAlaLeuValArgMetGly 677
 ||||| |||||
 131681 TTGTTTATTACGTTGACACAGAGCCCTGCCACGCTCTTGTGATAGTGGTCC 131730
 678 LysAspAlaThrLeuLeuHis..... 685
 ||||| |||||
 131731 AGAGATGCCACTGTACTAATACATAGGTAAGATGAACATAAATCTTTTG 131780
 685 685
 131781 GACAACATAATTCATTTATACCAACTTGATATCCAGTGCATTCATGG 131830
 685 685
 131831 TGCTATGATCTATATCTCCCTTTTGAGTGTTCCTCATCTGCTTA 131880
 685 685

OM of: US-09-434-382-2 to: N_Geneseq_36:* out_format : pfs
Date: Feb 18, 2001 9:25 PM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:
-MODEL=frame-p2n.model -DEV=xlpl
-Q=/cgnl_1/USPTO.spool/US09434382/runat_16022001_105749_29718/app_query.fasta_1.895
-DB=N_Geneseq_36 -QFMT=fastap -SUFFIX=ring -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FCGAPOP=6.000 -FCGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blomsum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=200000000
-USPR=US09434382_@CGN1_1_0 -NCPU=6 -LCPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1

Search information block:
Query: US-09-434-382-2
Query length: 826
Database: N_Geneseq_36:*
Database sequences: 480022
Database length: 187831343
Search time (sec): 105.780000

score_list:	Sequence	Strd Orig	ZScore	EScore Len	Documentation
	/cgnl_8/gcgdata/geneseq/NA2000.DAT:280231 +		404.00	613.88	238
	/cgnl_8/gcgdata/geneseq/NA1998.DAT:296286 -		221.50	305.90	3258
	/cgnl_8/gcgdata/geneseq/NA1998.DAT:V52183 -		217.50	294.27	5338
	/cgnl_8/gcgdata/geneseq/NA1999.DAT:X20534 +		195.00	242.57	25187
	/cgnl_8/gcgdata/geneseq/NA1999.DAT:X13218 +		187.50	247.47	5830
	/cgnl_8/gcgdata/geneseq/NA2000.DAT:A26852 +		175.50	238.63	2494
	/cgnl_8/gcgdata/geneseq/NA1999.DAT:X20248 +		172.50	168.09	910715
	/cgnl_8/gcgdata/geneseq/NA1999.DAT:Z22700 +		131.50	162.27	5660
	/cgnl_8/gcgdata/geneseq/NA1997.DAT:T58899 -		129.00	166.76	2692
	/cgnl_8/gcgdata/geneseq/NA1997.DAT:X30827 +		124.50	177.62	551
	/cgnl_8/gcgdata/geneseq/NA1998.DAT:T98612 +		123.50	165.46	1426
	/cgnl_8/gcgdata/geneseq/NA1998.DAT:X34766 +		123.00	143.20	9759
	/cgnl_8/gcgdata/geneseq/NA1995.DAT:Q90529 +		122.00	153.13	3501
	/cgnl_8/gcgdata/geneseq/NA1998.DAT:N80456 +		121.50	154.14	2987
	/cgnl_8/gcgdata/geneseq/NA1997.DAT:T59895 +		120.50	150.84	3501
	/cgnl_8/gcgdata/geneseq/NA1997.DAT:T89842 +		120.00	138.62	9759
	/cgnl_8/gcgdata/geneseq/NA2000.DAT:A08043 +		120.00	138.62	80161
	/cgnl_8/gcgdata/geneseq/NA1999.DAT:Z21501 +		119.00	113.55	19024
	/cgnl_8/gcgdata/geneseq/NA1999.DAT:X13116 +		118.50	128.87	50341
	/cgnl_8/gcgdata/geneseq/NA1998.DAT:V22674 +		116.50	114.93	50341
	/cgnl_8/gcgdata/geneseq/NA2000.DAT:Z39519 +		116.50	114.93	52297
	/cgnl_8/gcgdata/geneseq/NA1995.DAT:T51411 +		116.50	114.51	52297
	/cgnl_8/gcgdata/geneseq/NA1993.DAT:Q47357 +		116.50	131.74	9762
	/cgnl_8/gcgdata/geneseq/NA2000.DAT:Z48249 +		115.00	139.53	4541
	/cgnl_8/gcgdata/geneseq/NA1999.DAT:T72100 +		115.00	131.74	29879
	/cgnl_8/gcgdata/geneseq/NA1993.DAT:Q46806 +		115.00	118.47	9757
	/cgnl_8/gcgdata/geneseq/NA1995.DAT:Q97686 +		114.50	130.22	2339
	/cgnl_8/gcgdata/geneseq/NA1996.DAT:T35251 -		114.00	145.42	3109
	/cgnl_8/gcgdata/geneseq/NA1998.DAT:V34853 +		114.00	142.24	6400
	/cgnl_8/gcgdata/geneseq/NA1999.DAT:Z22701 +		114.00	134.17	2359
	/cgnl_8/gcgdata/geneseq/NA1996.DAT:T31273 +		113.50	144.56	5510
	/cgnl_8/gcgdata/geneseq/NA2000.DAT:Z51671 +		113.50	135.08	6847
	/cgnl_8/gcgdata/geneseq/NA1999.DAT:V99922 +		113.50	132.65	42235
	/cgnl_8/gcgdata/geneseq/NA1999.DAT:X98035 +		113.50	112.31	44377
	/cgnl_8/gcgdata/geneseq/NA1997.DAT:T78508 +		113.50	111.76	44377
	/cgnl_8/gcgdata/geneseq/NA1997.DAT:T80414 +		113.50	111.76	77536
	/cgnl_8/gcgdata/geneseq/NA2000.DAT:A14651 +		113.50	105.52	5970
	/cgnl_8/gcgdata/geneseq/NA2000.DAT:Z56003 +		113.00	133.42	1922
	/cgnl_8/gcgdata/geneseq/NA1994.DAT:Q66639 +		112.00	144.56	3279
	/cgnl_8/gcgdata/geneseq/NA1993.DAT:Q49397 +		112.00	138.59	

/cgnl_8/gcgdata/geneseq/NA2000.DAT:000334 - 111.50 139.43 5.61 2842
/cgnl_8/gcgdata/geneseq/NA1995.DAT:087426 + 111.00 141.51 4.29 2204
/cgnl_8/gcgdata/geneseq/NA1998.DAT:V30458 - 110.50 79.36 1.2e+04 53472
/cgnl_8/gcgdata/geneseq/NA1998.DAT:V30459 - 110.50 79.32 1.2e+04 53616
/cgnl_8/gcgdata/geneseq/NA1990.DAT:Q06470 - 110.00 136.22 8.46 3086
seq_name: /cgnl_8/gcgdata/geneseq/NA2000.DAT:280231
seq_documentation_block:
ID 280231 standard; CDNA; 238 BP.
AC 280231;
XX
DT 07-APR-2000 (first entry)
XX
DE Human colon cancer cell line SW480 cDNA clone SEQ ID NO:315.
XX
KW Human; gene expression product; diagnosis; tumour; colon cancer;
KW colorectal adenocarcinoma; cell line SW480; cell proliferation;
KW cytostatic; sarcoma; breast cancer; neoplasia; dysplasia;
KW hyperplasia; ds.
XX
OS Homo sapiens.
XX
PW W09964576-A2.
XX
PD 16-DEC-1999.
XX
PF 09-JUN-1999; 99WO-IB01062.
XX
PR 10-JUN-1998; 98US-0088801.
XX
PA (FARB) BAYER CORP.
XX
PI Endege WO, Steinmann KE, Astle JH, Burgess CC, Bushnell SE;
PI Carroll E, Catino TJ, Derti A, Ford DM, Lewis ME, Monahan JE;
PI Schlegel R;
XX
DR WPI; 2000-087220/07.
XX
PT Novel nucleic acids, used to develop products for the diagnosis and
PT treatment of disorders involving unwanted cell proliferation,
PT particularly cancers, especially colon cancer -
XX
Claim 15; Page 258; 469pp; English.
XX
CC Z79917 to 280766 represent double stranded cDNA clones isolated from the
CC human colorectal adenocarcinoma (colon cancer) cell line SW480. The
CC cDNA clones can be used to generate antisense oligonucleotides which
CC can be used for antisense therapy. Methods and products from the present
CC invention can be used for identifying and/or classifying cancerous cells
CC present in a human tumour, particularly in solid tumours, e.g. carcinomas
CC and sarcomas, e.g. breast or colon cancers. The cDNA clones can be used
CC for developing agents for the diagnosis and treatment of disorders
CC involving unwanted cell proliferation, such as neoplasia, dysplasia or
CC hyperplasia.
XX
SQ Sequence 238 BP; 55 A; 57 C; 69 G; 57 T; 0 other;

alignment_scores:
Quality: 404.00 Length: 78
Ratio: 5.179 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-434-382-2 x Z80231 ..
Align seg 1/1 to: 280231 from: 1 to: 238
60 LeuGlnValValAlaAlaGlySerArgAspSerGlyAlaAlaLeuTyrrVa 76
|||||
3 CTGACGGTGGTGGCGGGTACGGGGGACTCGGGGCGCGCTCTACGT 52

76 lPheSerGluPheAsnArgTyrLeuPheAsnCysGlyGluGlyValGlnA 93
53 CTTCTCCGAGTTCACCGGTATCTCTTCAACTGTGGAGAGCGCTTCAGA 102
93 rgLeuMetGlnGluHisLysLeuLysValAlaArgLeuAspAsnIlePhe 109
103 GACTCATGCGAGGACCAAGTTAAAGGTCTCGCTGGACCAACATATTC 152
110 LeuThrArgMetHisTrpSerAsnValGlyGlyLeuSerGlyMetIleLeu 126
153 CTGACACCAAGTCGACTGGTCTAATGTTGGGGGCTTAAGTGGAAATGATCT 202
126 uThrLeuLysGluThrGlyLeuProLysCysVal 137
203 TACTTTTAAAGGAAACCGGGCTTCCAAAGTCTGTA 236

seq_name: /cqn1_8/qcdata/geneseq/geneseq/NA1998.DAT:296286

seq_documentation_block:

ID Z96286 standard; DNA: 3258 BP.

AA 7.96286:

XX
XX
007067

DT 10-APR-

XX - XX

DE *S. pneumoniae* derived DNA from ORF #114.

XX
XX
XX

KW Treatment; prevention; disease; diagnosis; gene therapy; screening;

KW bacterial; antimicrobial; antibiotic; pathogenesis; infection; ss.

XXXX

OS Streptococcus

XX

PN WO9806734-A1.

XX

PD 19-FEB-1998.

XX
XX

PF 15-AUG-1997; 97WO-US14436.

XX
DB
16-AUG-1996
0808-0034033

PR 16-AUG-1996; 96US-0024022.
yy

XX
BA / SMIK \ SMITHKLINE BEECHAM CORP

PA (SMIK) SMITHKLINE BEECHAM CORP.
XX

XX
PT Black MT. Hodason J.E. Knowles D.JC. Lonetto MA. Nicholas RO.

PI Black Mt, 1
Stodola RK:

XX
XX
'NN NTONOS I I

DR WPI; 1998-159452/14.

DR P-PSDB; Y85922, Y859

XX
XX

PT Streptococcus pneumoniae protei

PT screening compounds for antibacterial activity

XX

PS Claim 4; Page 141-142; 640pp; English.

XX

CC This invention describes novel isolated Strept

CC polynucleotides (see Z96173-Z96494) and their encoded proteins (see
CC pY5792-Y8612). The DNA, vectors and host cells described in the method
CC of the invention are useful for the recombinant expression of the
CC polypeptides. The polypeptides are useful for treatment or prevention of
CC disease, or diagnosis of disease related to expression or activity of
CC such a polypeptide. They can also be used to screen for compounds which
CC interact with and inhibit or activate such a polypeptide. The
CC polypeptides (or DNA encoding them, via gene therapy) are also useful
CC for inducing an immunological response in a mammal. The antagonists are
CC useful to inhibit such bacterial polypeptides. The polypeptides are
CC particularly useful to identify antimicrobial compounds and antibiotics.
CC They are also useful to determine their role in pathogenesis of
CC infection, dysfunction and disease.

XX

SQ Sequence 3258 BP; 907 A; 723 C; 675 G; 953 T; 0 other;

alignment_scores:

```

Quality: 221.50      Length: 332
Ratio: 1.393        Gaps: 10
Percent Similarity: 47.892      Percent Identity: 24.398

alignment_block:
US-09-434-382-2 x 296286/rev ..

Align seg 1/1 to reverse of: Z96286 from: 1 to: 3258

482 GluIleIlePheLeuGlyThrGlySerAlaIleProMetLysIleArgAs 498
: : : : : | | | | | | | | | | | : : : : : | | | | |
1851 GATATTCAATTTTAGGAACGGGGCTGGTCAGCCCTCTAAAGCCGCGAA 1802
: : : : : | | | | | | | | | | | : : : : : | | | | |

498 nValSerAlaThrLeuValAsnIleSerProAspThrSer.....LeuL 513
| | | | | : : : : : | | | | | : : : : : | | | | |
1801 CGTTTCAAGTCTCGCCCTGAAACTTTTGGACGAGATTAACGAAGTTTGGC 1752
| | | | | : : : : : | | | | | | | | | | | | | | | |
1751 TCCTTGAGCTGTGGAGAAGGTACG.....: : : : : : : : : : 1729
: : : : : | | | | | | | | | | | : : : : : | | | | |

530 AspGlnValAspArgValLeuGlyThr.....LeuAl 540
: : : : : | | | | | | | | | | | : : : : : | | | | |
1728 .....CAAATCGCATTTCTGAAACCAACAATTCGACCACGTAAGTCA 1685
: : : : : | | | | | | | | | | | : : : : : | | | | |

540 aAlaValPheValSerHisLeuHisAlaAspHisHisThrGlyLeuPro 557
: : : : : | | | | | | | | | | | : : : : : | | | | |
1684 CAAATCTTTATTACCCATCTGCATGGAGACACACATTTTGGTTTGGCCAG 1635
: : : : : | | | | | | | | | | | : : : : : | | | | |

557 erIleLeuLeuGlnArgGluArgAlaLeuAlaSerLeuGlyLysProLeu 573
: : : : : | | | | | | | | | | | : : : : : | | | | |
1634 GTTTCCTTTCTAGCCGT.....GCCTTCAGCCCAATGAAGAGCAG 1594
: : : : : | | | | | | | | | | | : : : : : | | | | |

574 HisProLeuLeuValValAlaProAsnGlnLeuLysAlaTrpLeu.... 588
| | | : : : : : | | | | | : : : : : | | | | |
1593 ACAGATTGGAAATCTACGGACCTCAAGGAATCAAGTCATTTGCTTAA 1544
588 .....: : : : : : : : : : : : : : : : : : : : : 588

1543 CAGCCTTGTGTGTGTCAGGTTCTGCTCCCTACCGCATTCATTTCCATG 1494
589 .....: : : : : : : : : : : : : : : : : : : : : 592
1493 AGTTTGACCAAGATTCTCTGGTAAATTTCTGAAACCGATAAATTCAT 1444

593 AsnGlnCysGlnValLeuHisHisIleSerMetIleProAlaLysCys 609
: : : : : | | | | | | | | | | | : : : : : | | | | |
1443 GTGTATGCAGAGGAGTGGACACACTATTTTCTGTGTGGCTATCGTGT 1394

609 sLeuGlnIleGlyAlaGluIleSerSerProAlaValGluArgLeuIle 626
: : : : : | | | | | | | | | | | : : : : : | | | | |
1393 CATGCAAAAGGATCTAGAA.....: : : : : : : : : : : 1375

626 erSerLeuLeuArgThrCysAspLeuGluGluPheGlnThrCys..... 640
| | | | | : : : : : | | | | | : : : : : | | | | |
1374 .....GGACCGTGTGATGCTGAAAAAACAATCAAGGCTGCTGGTGT 1336

641 .....LeuValArgHisCysLysHisAlaPheGlyCysAlaLe 653
| | | : : : : : | | | | | : : : : : | | | | |
1335 CCGTTCGGCCCGCTTTTGGTAAAAATCAAAAAC.....GGCCAGGATCT 1292

653 uValHisThrSerGlyTrpLys.....: : : : : : : : : : 660
| | | | | : : : : : | | | | | : : : : : | | | | |

1291 TGTTTGGAGACCGGAAGTGAATCAAGGCAGACACTATATCTCAGCGC 1242

661 .....ValValTrpSerGlyAspThrMetProCysGlu 671
: : : : : | | | | | | | | | | | : : : : : | | | | |
1241 CACGTCCAGGTGAAGATTATCACTATTTTAGGAGACACTCGAAAAACGAT 1192

672 AlaLeuValArgMetGlyLysAspAlaThrLeuLeuIleHisGluAlaTh 688
| | | | | : : : : : | | | | | : : : : : | | | | |
1191 GCCAGTGTGCTCTGGCTGTCATGATGATGCTTATGTTAGTGTAGTCCAC 1142

```


661ValValTyrSerGlyAspThrMetProCysGlu 671
 2044 CACGTCACAGTAAAGATTATCACTATTTTGGAGACACTCGAAACGGGT 1995
 672 AlaLeuValArgMetGlyLysAspAlaThrLeuLeuIleHisGluAlaTh 688
 1994 GCCAGTGCCTGCTGGCTCAATGACAGATGCTTAGTTTCATGATCCAC 1945
 688 rLeuGluAspGlyLeuGluGluGluAlaValGluLysThrHisSerThrT 705
 1944 TTATGCAAGGGTGATGAAATAATTGCTGTAACCATGCTCACTCAACTA 1895
 705 hrSerClnAlaIleSerValGlyMetArgMetAsnAlaGluPheIleMet 721
 1894 ATATGCAAGCTGCACAAGTAGCGGTAGAGAGCGGTGCCAAGCGCTCCTA 1845
 722 LeuAsnHisPheSerGlnArgTyr.....AlaLysValPr 733
 1844 CTCACCATATCAGTCCCGCTTCTCTCAAAAGATATTAGCAAACTCAA 1795
 733 oLeuPheSerProAsnPheSerGlyLysValGlyValAlaLapheAsp 748
 1794 GAAAGACGCTGCCACAATTTTGAATAATGTCATGTGGTCAAGAC 1749

seq_name: /cgnl_8/gcdata/geneseq/geneseqn/NA1999.DAT.X20534

seq_documentation_block:

XX X20534 standard; DNA; 25187 BP.

AC X20534;

DT 05-MAY-1999 (first entry)

DE Polynucleotide sequence from the genome of *Treponema pallidum*.

XX *Treponema pallidum* infection; syphilis; Borrelia infection; animal;
 KW enzyme production; ds.

XX *Treponema pallidum*.

OS WO9859034-A2.

XX 30-DEC-1998.

XX 23-JUN-1998; 98WO-US13041.

XX 24-JUN-1997; 97US-0050667.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Fraser CM;

PI WPI; 1999-081273/07.

XX New isolated *Treponema pallidum* nucleic acids - used to develop
 PT products for the detection, diagnosis, characterisation, prevention
 PT and therapy of *T. pallidum* infections, particularly syphilis

PS Claim 1; Page 375-389; 1150pp; English.

XX X20500-21243 represent polynucleotide sequences from the genome of
 CC *Treponema pallidum*. The sequences can be used for detection,
 CC diagnosis, characterisation, prevention and therapy for *T. pallidum*
 CC infections, particularly syphilis. They can also be used for detecting
 CC diseases related to *Borrelia* infections in animals, and for the
 CC production of biosynthetic products such as enzymes.

XX Sequence 25187 BP; 6051 A; 7307 C; 6131 G; 5671 T; 27 other;

alignment_scores:

Quality: 195.00

Ratio: 1.010

Length: 393

Gaps: 13

Percent Similarity: 49.109 Percent Identity: 21.628
 alignment_block:
 US-09-434-382-2 x X20534/rev ..
 Align seg 1/1 to reverse of: X20534 from: 1 to: 25187
 395 AspIlePheProLeuLeuThrSerPheArgCysLysLysGluGlyProTh 411
 14265 GAGATCTTTCCGCTCACCAAGCGTTCGGAAGGACAGACATCGCGTACT 14216
 411 rLeuSerValProMetValGlnGlyCysLeuLeuLysTyrGlnLeuA 428
 14215 AACAGCGGTGAATTTTGAACCGTCGAATCGCTATGCTGCTATCGCCTG 14166
 428 rgProArgArgGluTrpGlnArgAspAlaIleIleThrCysAsnProGlu 444
 14165 GACTTCCGCGCCAGTACGACGAGAC.....AAA 14137
 445 GluPheIleValGluAlaLeuGln.....LeuProAsnPh 456
 14136 GGCTTTGCACTTGTATGATACCGCGTGGTTCGCTTTCGCTCTGAGTA 14087
 456 eGlnGlnSerValGlnGluTyrArgArgSerAlaGlnAspGlyProAlap 473
 14086 CAAAGAGCACAAAGATCCATCAGCGCCTGAGT.....TAACCTCCAA 14043
 473 roAlaGluLysArgSer.....GlnTyrProGluIleIlePheLeu 486
 14042 AAGGTTCAGTCGGAATGGTGCATATGAATCTTGAACGGTTTATTTTA 13993
 487 GlyThrGlySerAlaIlePrometLysIleArgAsnValSerAlaThrLe 503
 13992 GGCTGGCGGCATGGTGCATTACCCCATCGCATCTGACTTCTGTGTT 13943
 503 uValAsnIleSerProAspThrSerLeuLeuLeuAspCysGlyGluGlyT 520
 13942 GTTGGCTGCTGAGGAGAGTTC...TTTCTGTTCACGCTGGGGAAGSTA 13896
 520 hrPheGlyGlnLeuCysArg...HisTyrGlyAspGlnValAspArgVal 535
 13895 CCCAAGTGTCTGGCGCTCTGAGTTACGCTGGAAGAAA..... 13856
 536 LeuGlyThrLeuAlaAlaValPheValSerHisLeuHisAlaAspHisH 552
 13855ATCAGTGCATTTTCATCAGTCACCCCATGCAGACACAT 13815
 552 sThrGlyLeuProSerIleLeuLeuGlnArgGluArgAlaLeuAlaSerL 569
 13814 CACTGGCTGCGCGGCTTTTGTATGCTCTCTTCTCAAGTTGCTCGCAGC. 13766
 569 euGlyLysProLeuHisProLeuValValAlaProAsnGlnLeuLys 585
 13765GAACGGCTGTATCATCTCGTCTCCCAAGAACTGCA 13730
 586 AlaTrpLeuGlnGlnTyrHisAsnGlnCysGlnGluValLeuHisHisI 602
 13729 GAGTATGTGGAACAGCGCAGGATTTTGGACATGTACATTAATACGA 13680
 602 eSerMetIleProAlaLysCysLeuGlnGluGlyAlaGluIleSerSerP 619
 13679 GATTATTGTC.....AAAGAGGTGATAGAAC 13654
 619 roAlaValGluArgLeuIleSerSerLeuLeuArgThrCysAspLeuGlu 635
 13653 CACAAGTG.....GTATACCGTGGTAAGGATTTTCAG 13622
 636 GluPheGlnThrCysLeuValArgHisCysLysHisAlaPheGlyCysAl 652
 13621 GTGCGCTGTTTGTCTG...GATCATACCAAGCGGTGTATGGGTACAC 13575
 652 aLeu..... 653


```

903 GACAGGAATTAATGGACACGGTTGTGGAGACCGTAAACCTGGACGA 952
660 LysValValTyrSerGlyAspThrMetProCysGluAlaLeuValArgMe 676
953 ATCGTGACGATTTTAGGAGACACTCGCAAAACCAAAATAGTGAACCTT 1002
676 tGlyLysAspAlaThrLeuLeuIleHisGluAlaThrLeuGluAspGlyL 693
1003 AGCAGCAGCGGATGTTTAGTCCATGAAAGTACGTTTAATAAACACG 1052
693 euGluGluAlaValGluLysThrHisSerThrThrSerGlnAlaIle 709
1053 AAGCAAAATGGCAAAACCTTCTCCACTCAACGAGTCACACAGCACT 1102
710 SerValGlyMetArgMetAsnAlaGluPheIleMetLeuAsnHisPheSe 726
1103 GAAGTGGCAAAAGAGCCCAAGTAAACAACATTTATTAACGCATATTAG 1152
726 rGlnArgTyr 729
1153 TGCCCGTTAT 1162
```

seq_name: /cgnl_8/gcgdata/geneseq/NA2000.DAT.A26852

seq_documentation_block:
ID A26852 standard; DNA; 2494 BP.

XX AC A26852;

XX DT 29-JUN-2000 (first entry)

XX DE Essential Staphylococcus aureus gene #3.

XX KW Essential gene; Staphylococcus aureus infection; screening; prevention;
XX KW antibacterial agent; bacterial infection; ss.

XX OS Staphylococcus aureus.

XX PN US6037123-A.

XX PD 14-MAR-2000.

XX PF 13-SEP-1996; 96US-0714918.

XX PR 15-SEP-1995; 95US-0003798.

XX PR 22-DEC-1995; 95US-0009102.

XX PA (MICR-) MICROCID PHARM INC.

XX PI Benton B, Lee VJ, Martin PK, Schmid MB, Sun D, Malouin F;

XX DR WPI; 2000-282222/24.

XX PT Novel methods for screening for antibacterial agents, useful for
PT treatment or prevention of Staphylococcus infection, by testing
PT compounds against, essential bacterial genes -

XX PS Disclosure; Column 35-38; 256pp; English.

XX CC The present sequence represents an essential Staphylococcus aureus gene.
CC The invention relates to a method for screening for an antibacterial
CC agent. The method comprises determining if a test compound is active
CC against the essential bacterial genes (A26850-A26956). The invention also
CC includes a method for evaluating an agent for activity on the essential
CC genes. Also included in the invention is the production of an
CC antibacterial agent. The antibacterial agents of the invention are used
CC to treat or prevent bacterial infections, particularly where caused by
CC Staphylococcus aureus. The antibacterial agents are unlikely to be
CC affected by known resistance mechanisms.

XX SQ Sequence 2494 BP; 809 A; 396 C; 389 G; 880 T; 20 other;

alignment_scores:
Quality: 175.50 Length: 120
Ratio: 2.279 Gaps: 3
Percent Similarity: 64.167 Percent Identity: 34.167

alignment_block:
US-09-434-382-2 x A26852 ..
Align seg 1/1 to: A26852 from: 1 to: 2494

```

476 LysArgSerGlnTyrProGluIleIlePheLeuGlyThrGlySerAlaIle 492
2154 AAAGAGGGGCATATATGCAAGTTTACATTTTGGACGAGTGCAGGTTT 2203
492 eProMetLysIleArgAsnValSerAlaThrLeuValAsnIleSerPro. 508
2204 GCCTACAAAGAGAGAAATACACAGCAATCGCCTTAATTTAGAACCAT 2253
509 .....AspThrSerLeuLeuLeuAspCysGlyGluGlyThrPheGlyGln 523
2254 ATTCCAATTCCATATGCGCTTTTCGACGTTGTGAAGGTACACAGCACCAA 2303
524 LeucysArgHisTyrGlyAspGlnValAspArgValLeuGlyThrLeuAl 540
2304 ATTTTACATCAT.....GCAATTAAATTTAGGAAAAGTGAC 2338
540 aAlaValPheValSerHisLeuHisAlaAspHisHisThrGlyLeuProS 557
2339 ACATATATTTATTACTCATATGCGATGCGATCATATTTTGGTTGCCAG 2388
557 erIleLeuGlnArgGluAlaLeuAlaSerLeuGlyLysProLeu 573
2389 GATTACTTTCTAGTCGT.....TCTTTTCAGGCGGTGAACAG 2426
574 HisProLeuValValAlaProAsnGlnLeuLysAlaTrrpLeuGlnGln 590
2427 AAGCGCTTACATTTGGTTGGACCAAAAGGAATTAAGCATATATGTGGAAT 2476
590 nTyrHisAsn 593
2477 GTCTATGAAT 2486
```

seq_name: /cgnl_8/gcgdata/geneseq/NA1999.DAT.X20248

seq_documentation_block:

ID X20248 standard; DNA; 910715 BP.

XX AC X20248;

XX DT 04-MAY-1999 (first entry)

XX DE Borrelia burgdorferi polynucleotide sequence #1.

XX KW Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;
KW epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;
KW infection; diagnosis; characterisation; detection; ds.

XX OS Borrelia burgdorferi.

XX PN W09858943-A1.

XX PD 30-DEC-1998.

XX PF 18-JUN-1998; 98WO-US12764.

XX PR 03-SEP-1997; 97US-0057483.

XX PR 20-JUN-1997; 97US-0050359.

XX PR 22-JUL-1997; 97US-0053344.

XX PR 22-JUL-1997; 97US-0053377.

XX PA (HUMA-) HUMAN GENOME SCI INC.

PA (MEDI-) MEDIMMUNE INC.


```

112095 .....TATGAAGATAAAAAACAAAAAATTGA 112120
643 gHis.....CysLysHisAlaPheGlyCysAla..... 652
::: |||||::: ||:::
112121 ATACACTAACTAAAACATTCATAGAATCGTTGGATATTTATTATTAG 112170
652 ..... 652
112171 AAAAAGATAAACCCGGCAAATTCAACACAGAAAAAGCAGAAGAGCTAAAT 112220
653 .....Le 653
112221 ATTCTAAAGGCCCTATTAGAAAAGCCCTACAAGATGGAAAAAGAAATATT 112270
653 uValHis.....ThrS 657
|||||:
112271 GGTAAACGAAAAANTATAAAGCCATCAGAAATACTTGGAAAAATCTAAAA 112320
657 erGlyTrpLysValValTySerGlyAspThrMetProCysGluAlaLeu 673
::: |||||::: ||:::
112321 AAGGACTAAAGCTTGCATACATTACAGATACTGGTTATTTTAAGAACTC 112370
674 ValArgMetGlyLysAspAlaThrLeuLeuIleHisGluAlaThrLeuGl 690
:::::::::: ||::: |||||:::
112371 ATACAGCAAAATCAAAAATTTTAACCTTGTAATAATTGAGAGCACATTAA 112420
690 uaspGlyLeuGluGluAlaValGluLysThrHisSerThrThrSerG 707
::: |||||::: ||||| ||||| |||||:::
112421 AAATGAGCTAAAAGAGAGCCGATAAAAAACTTCATTAACAGCTGGCG 112470
707 InAlaIleSerValGlyMetArgMetAsnAlaGluPheIleMetLeuAsn 723
||| ::::: ::::: |||
112471 GGGCTGCAAAATATTGTCAGCAAGCAAAAGCTTTTACAAACAGGACTATC 112520
724 HisPheSerGlnArgTyr 729
|||||::: |||||
112521 CATTTTAGTGAAGATAT 112538

seq_name: /cgnl_8/gcgdata/geneseq/geneseqn/NA1999.DAT:222700
seq_documentation_block:
ID 222700 standard; DNA; 5660 BP.
XX AC 222700;
XX DT 15-FEB-2000 (first entry)
XX DE Pseudomonas fluorescens ABC transporter cassette operon I.
XX KW ATP-binding cassette; transporter; operon; LipBCD; Serratia m
XX KW microbe; protein secretion; ds.
XX OS Pseudomonas fluorescens.
XX Key Location/Qualifiers
FH 778..2529
FT CDS /*tag= a
FT FT /product= "component A"
FT 2526..3860 /*tag= b
FT FT /product= "component B"
FT 3863..5197 /*tag= c
FT FT /product= "component C"
XX JF11276172-A.
XX PN
XX PD 12-OCT-1999.
XX PF 27-MAR-1998; 98JP-0080597.
XX PR 27-MAR-1998; 98JP-0080597.
XX XX

```

PA (TANA) TANABE SEIYAKU CO.
 XX WPI; 1999-626936/54.
 DR P-PSDB; Y55919, Y55920, Y55921.
 XX
 XX An ABC transporter gene - used to increase the ability of a microbe to
 PT secrete proteins
 XX
 XX Claim 4; Page 12-14; 28pp; Japanese.
 PS
 CC This sequence represents an ATP-binding cassette (ABC) transporter
 CC operon from *Pseudomonas fluorescens* strain 33. The operon comprises
 CC 3 genes where the termination codon of the first gene overlaps with
 CC the initiation codon of the second gene. The sequences of the encoded
 CC proteins have 60, 44 and 46% homology respectively to the LipBD
 CC proteins from *Serratia marcescens*. The novel gene and protein can
 CC give or increase the ability of a microbe for secreting a protein.
 XX
 XX Sequence 5660 BP; 1131 A; 1817 C; 1736 G; 976 T; 0 other;
 SQ

alignment_scores:
 Quality: 131.50 Length: 667
 Ratio: 0.431 Caps: 41
 Percent Similarity: 45.727 Percent Identity: 22.789

alignment_block:
 US-09-434-382-2 x 222700 ..
 Align seg 1/1 to: 222700 from: 1 to: 5660

10 SerAlaAlaGlyArgThrMetSerGlnGlyArgThrIleSerGlnAlaPro 26
 |||:|||||
 2082 AGTGTGCGCGCGC.....CC 2098

26 oAlaArgArgGluArgProArgLysAspPro.....LeuArgHisLeuA 41
 |||:|||||
 2099 AGCAAGCCGG...CGTGACCAACTGATCCTCGCTTGTGCCACCGGCTAC 2145

41 rGThrArgGluLysArg.GlyProSerGlyCysSerGlyGlyProAsnTh 57
 |||:|||||
 2146 GACACCGTGTCTCGGCGCAAGCAGCGCGCTGTCGGTGGCCCAACAA 2195

57 rValTyrLeuGlnValAlaAlaGlySerArgAspSerGlyAlaAlaL 74
 |||:|||||
 2196 G.....CGGTCGCGCTGCTCGCGCTGTATGGCGCGCGCGC 2236

74 euTyrValPheSerGluPheAsnArgTyrLeuPheAsnCysGlyGluGly 90
 |||:|||||
 2237 TCATCGTACTGATGAGCCCACTCCAACTCGACACCGTGGCGGAGGCC 2286

91ValGlnArgLeuMetGlnGluHisLysLeuLysValAlaAla 105
 |||:|||||
 2287 GCCCTGGCCGCGCATATGCAG.....ATGAAGGCCAGGCGCAG 2327

105 uAspAsnIlePheLeuThrArgMetHisTrpSerAsnValGlyGlyLeu 122
 |||:|||||
 2328 CACGCTGTGTGTGTGAGC.....CACCGCTCTCGCGCTGGCCAGG 2371

122 erGlyMetIleLeuThrLeuLysGluThrGlyLeuProLysCysValLeu 138
 |||:|||||
 2372 CCGACAAGCTGTGTGTCTCAACGAAGCCCGCTTGCAG.....GCC 2412

139 SerGlyProProGlnLeuGluLysTyrLeuGluAlaIleLysIlePheSe 155
 |||:|||||
 2413 TTCGACCCAGCCAG.....GACGTGCTCGCGGCACTGTC 2447

155 rGly.ProLeuLysGlyIleGluLeuAlaValArgProHisSerAlaPro 171
 |||:|||||
 2448 CGGCAGCAGGAAG..... 2461

172 GluTyrGluAspGluThrMetThrValTyrGlnIleProIleHisSerG 188
 |||:|||||

2462CACGGCGA 2469

188 uGlnArg.ArgGlyLysHisGlnProTrpGlnSerProGluArgProLeu 204
 |||:|||||
 2470 GAAAGCGCGGTGCTCAGCTCAGCGCTCAGTATCAAGCCGGAGGAATCC 2519

205 SerArgLeuSerProGluArgSerSerAspSerGluSerAsnGluAsnG 221
 |||:|||||
 2520 AGGCGCATGAGCAGCAGCAGCATTCAAATCAACCGCATACGAACA 2569

221 uProHisLeuProHisGlyValSerGlnArgArgGlyValArgAspSerS 238
 |||:|||||
 2570 GACTCACCCCAACATGGCGCA.....CGTTCTTTCGCCCGCATGGGT 2613

238 erLeuValValAlaPheIleCysLysLeuHisLeuLysArgGlyAsnPh 254
 |||:|||||
 2614 GCTCTCTGACGGTG.....GTGGCGCGCGTGTGTTTC 2645

255 LeuVal.....LeuLysAlaLysGluMetGlyLeuProVal...G 267
 |||:|||||
 2646 TTCCTGTGGCGCAGCTGGCGCGCTGGACAGGCGATTCCCGTGCAGG 2695

267 yThrAlaAlaIle.....AlaProI 274
 |||:|||||
 2696 CACCTGTGTGTCTCGGCAAGCGCGCTGCAAAACCCCTGAGCCCGC 2745

274 leIleAlaAla.....ValLysAspGlyLysSerIleThrHisGlu 287
 |||:|||||
 2746 GAGTGTCTACGCGGATCTGTGTGGGGAAGCGAGGCGAGTG...AAGCAG 2792

288 GlyArgGluIleLeuAlaGluGlu.....LeuCysT 298
 |||:|||||
 2793 GSCCAGCGCTGTTCGCGCTCGACCAAGCAAAACAGGCTGACGTGCA 2842

298 hrPro.....ProAspPro.....Gly 303
 |||:|||||
 2843 CTCCTGCAAGCCAGTACCGCATCGCTGGCGCCAGCCGTGGCGGTGC 2892

304 AlaAlaPheValValGluCysProAspGluSerPheIleGlnProI 320
 |||:|||||
 2893 AGAGCGCGGTGATACCACT..... 2913

320 eCysGluAsnAlaThrPheGlnArg.TyrGlnGlyLysAlaAspAlaPr 336
 |||:|||||
 2914 ...CGAGCATCACCTTCCGGCGAGCTGAGCGGTAAACCGGATCAAGC 2959

336 oValAlaLeuValValHisMetAlaProAlaSerValLeuValAspSerA 353
 |||:|||||
 2960 CTTGCCCTTGGTT.....CTGGAAGGCCAGCGCAACTGTTCAGCAGCC 3003

353 rGTYrGlnGlnTrpMetGluArgPheGlyProAspThrGlnHisLeuVal 369
 |||:|||||
 3004 GC.....CGCAAGCCTTTGCCGGGAACAGCGCGGGATTGCG 3041

370 LeuAsnGluAsnCysAlaSerValHis..... 378
 |||:|||||
 3042 CGAATCATCGAGGCGCCACCGCCCACTCAACGATCGCGCGCGCGCG 3091

379AsnLeuArgSerHisLysIleGlnThrGlnLeuAsnLeuI 392
 |||:|||||
 3092 CAGTACCTGCTGCCAGGCCCAATCTCTCGCCACCACTGTAACAAC 3141

392 leHisProAspIlePheProLeuLeuThrSerPheArgCysLysLysGlu 408
 |||:|||||
 3142 TGCAA..... 3146

409 GlyProThrLeuSerValProMetValGlnGlyGluCysLeuLeuLys 425
 |||:|||||
 3147 ...CCCTTGGCGCAGACAGCGGTATACATCCGCGCAACCGGCTGATGAGTA 3193

425 rGlnLeuArgProArgGluTrpGlnArgAspAlaIleThrCysA 442
 |||:|||||
 3194 CCAG...CGGCGAGCTGTCCAGGTGCAACAGGACCTGGCGCAACACCG 3240

[illegible]

alignment_scores:

Quality: 123.00 Length: 939
 Ratio: 0.330 Gaps: 48
 Percent Similarity: 39.723 Percent Identity: 19.276

alignment_block:

US-09-434-382-2 x V34766 ..

Align seg 1/1 to: V34766 from: 1 to: 9759

```

2 TrpAlaLeuCys.....SerLeuLeuArgSe 10
||| |||||
3943 TGCACCTGTGCGCGCAGCAGGAGGGGTGGCCCGAGGTACTACGAGC 3992

10 rAlaAlaGly.ArgThrMetSerGlnGlyArgThrIleSerGlnAlaPro 26
||| |||
3993 ACCTCGAGGTGCGCGCTCGGGATGACCCCATGGCCCGGGCGGCCTC 4042

27 AlaArgGluArgProArgLysAspProLeuArgHisLeuArgThrAr 43
|||
4043 GCATCAGTCCACAGCCCTCGCAAA..... 4066

43 gGluLysArgGlyProSerGlyCysSerGlyGlyProAsnThrValTyrL 60
|||||
4067 .....GGCCCTTACAATATCAGGG 4085

60 euGlnValValAlaAlaGlySerArgAspSerGlyAlaAlaLeuTyrVal 76
::
4086 TATGGACATGCGCGCAGGC.....CTGGCAAGACCACCCCGCATC 4126

77 PheSerGluPheAsnArgTyr...LeuPheAsnCysGlyGluGlyValG1 92
:::
4127 CTGCTGCTTACGCGCGAAGACCTTACGTCTGC.....CCCACAA 4170

92 nArgLeuMetGlnGluHisLysLeuLysValAlaArgLeuAspAsnIleP 109
:::
4171 TGGCTCTCTCAGCAGATCCAGCCCAAACTCCGCGCGCGCAT..... 4213

109 heLeuThrArgMetHisTrpSerAsnValGlyGlyLeuSerGlyMetIle 125
4213 ..... 4213

126 LeuThrLeuLysGluThrGlyLeuProLysCysValLeuSerGlyProPr 142
::
4214 ATCGAGATCAAGAAGCGCGCCACCTACGAGCGCGCTGACGAAACCGCT 4263

142 oGlnLeuLys.....TyrLeuGluAlaIleLysIlePheSerGly. 156
::
4264 CGCGCGCTACCGCGCATCTACATCATGATGAGGGTTTCACTCTCGGGGCG 4313

157 .....ProLeuLysGlyIleGluLeu 163
4314 AGTACTGCGGCTTCGTTGCCAGCAACACCGCGGAGGTGATCTGCGTC 4363

164 AlaValArgProHisSerAlaProGluTyrGluAspGluThrMetThrVa 180
::
4364 GGTGATCGGGACCAAGTCGCGCCACACTACGCCAATAACTGCCGCACC... 4411

180 lTyrGlnIleProIleHisSerGluGlnArgArgGlyLysHisGlnProt 197
::
4412 ....CCGTCCTCGACCGCTGGCTACCGAGCGCTCGCGCAC...ACTT 4454

197 rpGlnSerProGluArgProLeuSerArgLeuSerProGluArgSerSer 213
|||
4455 GGCCTTCCCGACTGCTGGCGGCGCGCTCGCGCGGGGCTCGATTAT 4504

214 AspSerGluSerAsnGluAsnGluProHisLeuProHisGlyValSerG1 230
|||
4505 GACATCGAG.....GGCGA 4518

230 nArgArgGlyVal.....ArgAspSerSerL 239

```

```

4519 GCGCACCAGCACCTTCGCTGCAACCTTTGGGACGGCCCGAGTGCACC 4568
||| |||
239 euValValAlaPhe.....IleCysLysLeuHisLeuLysArg 251
||
4569 TTCACCTCGCTTCTCGCGGAAACCGTGGCGCGCTTAC....GAGGCT 4615

252 GlyAsnPheLeuValLeuLysAlaLysGluMetGlyLeuProValGlyTh 268
|||
4616 GGCATACGCGCATACACCGTGGCGAGGCCAGGTATGAGCGTCGGCAC 4665

268 rAlaAlaIle..... 271
|||||
4666 CGCCTGCATCCATGTAGGACAGACGACCGAGTTCCTCGCGCTGA 4715

272 .....AlaProIle 274
|||
4716 CACGCGACCTCGCATCGTCAGCCTGACCCGGGCTCCGAGCACACTAC 4765

275 IleAlaAlaValLysAspGly..... 281
::
4766 CTCACGAGCTCGAGGACGCTCACTGCGCGCTGCGGGGCTCAGCGGTT 4815

282 .....LysSerIleThrHisGluGlyA 289
|||||
4816 CCTCGACGCGCGGCACCTGGCGGAGCTCAAGGAGTTTCCCGCTGGCATTC 4865

289 rgGluIleLeuAlaGluLeuCysThrProProAspProGlyAlaAla 305
::
4866 ACCGCGTGTGCGCGTCGACGAGCAGCCACCGTTCGCGCGCGCGCAC 4915

306 PheValValValGluCysProAspGluSerPheIleGlnProIleCysG1 322
::
4916 GGCATC.....CCGAGGCCAACAGCTGCGCGCTTCTGCC 4953

322 uAsnAlaThrPheGlnArgTyrGlnGlyLysAlaAspAlaProValAlaL 339
::
4954 CCGCACTCTGGAGGAGCTCTCTTCGCGCGTGGCGGCCACCCC..... 4996

339 euValValHisMetAlaProAlaSerValLeuValAspSerArgTyrGln 355
|||
4997 .....CATTACGCGGACCTCAACCGCTGACTGAGGCGGACAGAA 5038

356 GlnTrpMetGluArgPheGlyProAspThrGlnHisLeuValLeuAsnG1 372
|||
5039 .....GTGCGGTATATGCGCATCTCGCGTCACTG...CTCAACAA 5076

372 uAsnCysAlaSerValHisAsnLeuArgSerHisLysIleGlnThrGlnL 389
|||
5077 GAATCACACCGAGATGCCGA.....ACGGAACGCGTTC 5111

389 euAsnLeuIleHisProAspIlePheProLeuLeuThrSerPheArgCys 405
|||
5112 TCAGTGGCGTTTCGCCG.....TCGCGCTACCGCGCG 5143

406 LysLysGluGlyProThrLeuSerValProMetValGlnGlyGluCysLe 422
::
5144 GCGAGGATGGTTCGACCTCCGCACTGCTGGCGCCCGCAGCACCGCG 5193

422 uLeuLysTyrGlnLeuArgProArgArg.....G 432
|||
5194 CCCTTTTCGCGCAGATCCACCGCGCGCTCACTGCTGGGGTTCGCCCAGG 5243

432 luTrpGlnArgAspAlaIleIleThrCysAsnProGluGluPheIleVal 448
|||
5244 AGTGGCGCATG.....ACGTACTTGGCG 5266

449 GluAlaLeuGlnLeuProAsnPheGlnInSerValGlnLutyrArgAr 465
|||
5267 GAACGATCGACCTCACTGACGTCTACACGAGATGGCGTGGCGCGCG 5316

465 gSerAlaGlnAspGlyProAlaGluLysArgSerGlnTyrProG 482
::

```


Quality: 122.00 Length: 970
 Ratio: 0.998 Gaps: 48
 Percent Similarity: 42.165 Percent Identity: 19.175
 alignment_block:
 us-09-434-382-2 x 090529
 Align seg 1/1 to: 090529 from: 1 to: 3501

```

13 GlyArgThrMetSerGlnGlyArgThrIleSerGlnAlaProAlaArgAr 29
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
14 GCGGAGTTCCGCGCGAGGAGCCCTACGCGAGCCAGTTCGCGGAGGAG 63
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
29 gGluArg.....ProArgLysAspProLeuArgHisLeuArg..ThrArg 43
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
64 GCGCGCCAGTGCAGCGATGCGCGGCGAGTCCGCGCTCCAAGTTGTGGA 113
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
44 GluLysArgGlyProSerGlyCysSerGlyGlyProAsnThrValTyrLe 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
114 GAAGCTGCGGCGCGCTGCTGCGTCCGAAACCGGACCCCAA..... 152
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
60 uGlnValValAlaAlaGlySerArgAspSerGlyAlaAlaLeuTyrValP 77
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
153 .....GAAGCTATTGAAATAT 168
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
77 heSerGluPheAsnArgTyrLeuPhe.....AsnCysGlyGlu 89
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
169 TTGAAGAAACTTTCCGCTCTTACCTATTACAGTAGACATTCTTGTGGAG 218
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
90 .GlyValGlnArgLeuMetGlnGluHisLysLeuLysValAlaArgLeuA 106
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
219 TGGGGTGGGAAACAGTG..... 237
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
106 sPAsnIlePheLeu..ThrArgMetHisTrpSerAsnValGlyGlyLeuSe 122
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
238 ..AACAGCTTCCGAACAGGACCACTGGAA..... 268
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
122 rGlyMetIleLeuThrLeuLysGluThrGlyLeuProLysCysValLeuS 139
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
269 .....ACTTTCAGAGACCTGGTTGCC.....A 293
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
139 erGlyProProGlnLeu.....GluLysTyrLeuGluAlaIleLys 152
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
294 GTGGAAGAAGCTGTTCCAGTAGAAGCAAAATGAGGCTGAGGATCAGG 343
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
153 IlePheSerGlyProLeuLysGlyIleGluLeuAlaValArgProHisSe 169
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
344 ATTTTGAGAGACCAATTCCCGAAG.....CGTCCCGGAGA 380
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
169 rAlaProGluTyrGluAspGluThrMetThrValTyrGln..... 182
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
381 TGTTCGCCAGGAGGAGGAGGCGGAGGCGGAACTACCAGGAAGCTGGC 430
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
183 .....IleProIleHisSerGluGlnArgArgGlyLysHis 194
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
431 AAGCCCTCAGCACCCAGCCCTACAGCCCTGAGCACACAGACAGAAAAAG 480
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
195 GlnProTrpGlnSerProGluArgPro..... 203
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
481 AGAAACTTCCTGAGCTTGAAGAGCCCTCACAAGTGGCTCATGTCACGA 530
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
204 .....LeuSerArgLeuSerProGluArgSers 213
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
531 GAGGAGAGATGAAAGGAGGAGGTGTCAAAAGTGTACACCATATTCTT 580
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
213 erAspSerGluSerAsnGlu.....Asn 220
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
581 CAGACCCCGAGTCTCTGAGCTATGCTATGCTATGCTCAATCTCTCCACC 630
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
221 GluProHisLeuProHisGlyValSerGlnArgArgGlyValArgaspSe 237
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
631 AGTCCCATCAATGTATACAGACCTCTCTAGGTCCCGGAGATGGACCA 680
  
```

```

237 rSerLeuValValAlaPheIleCysLysLeuHisLeuLysArgGlyAsn 254
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
681 GGAACCCATCGTCTCA.....CACCCGAGGCTGGGAAG 715
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
254 heLeuVal.....LeuLysAlaLysGluMetGlyLeu 264
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
716 TCCACAGTAATACCTTTCAGGACAGACTAGGGGTAGTCACCTGGGTGAG 765
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
265 ProValGlyThrAlaAlaIleAla..... 272
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
766 CACCAAGGGAAGGGCTGTAGCCAAACCAAGCCACACAAATCTTCCCA 815
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
273 .....ProIleIleAlaValLysAspGlyLysSerIle. 284
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
816 CAAGGAGAAACGCCCTGTGATGCCAGGGGATGAGAAGAGCTCTGTCA 865
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
285 .....ThrHisGluGlyArgGlu 290
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
866 TGGCGCAGAGAGATGCACAAAGCCCTTTCCAAAGAGGAGCGCGAAGG 915
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
291 IleLeuAlaGluGluLeuCysThrProProAspProGlyAlaAlaPheVa 307
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
916 CTACTCTCAGAGGACAGTGCACAGGAGAAGCTGCCCTCCAGT.....GT 959
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
307 lValValGluCysProAspGluSerPheIleGlnProIleCysGluAsnA 324
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
960 TGTCAAGAAAGAGAGACAGA.....AGCCTCAAG 999
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
324 laThrPheGlnArgTyrGlnGlyLysAlaAspAlaProValAlaLeuVal 340
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
982 .....GAAGGCAAC.....AGCCTCAAG 999
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
341 ValHisMetAlaProAla...SerValLeuValAspSerArgTyrGlnG 356
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
1000 AAGAGTGTGCACCTGCCTTAGATGTGCTTCAGACCAACACCTTTAAAA 1049
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
356 nTrpMetGluArgPheGlyProAspThrGlnHisLeuValLeuAsnGluA 373
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
1050 GCCCAACACAAAG.....GACTCCGAGAAATCAAGTCTGACAAA 1090
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
373 snCysAlaSerVal..... 377
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
1091 AACAGCAGAGTGTAGATAGCTGAGCTCAGACGAGGACAGGACCCA 1140
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
378 .....HisAsnLeuArgSerHis..... 383
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
1141 TTACCCAGAGCCAGGATAAGTTCCCAACACCTGAGGCTCAGGAGG 1190
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
384 .LysIleGlnThrGlnLeuAsnLeuIleHisProAspIlePheProLeuL 400
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
1191 GAAAGTAAGAACTAACTCGGATCGAAAGTCACCAGGCTCACTCCCTAAAG 1240
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
400 euThrSerPheArgCys.....LysLysGluGlyProThrLeuSerVal 414
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
1241 TAGAAGAGATGGACATGGATCATGAGTTGAGCAGGCCACCACCATGCT 1290
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
415 ProMetValGlnGlyGluCysLeuLysTyrGlnLeuArgProArgAr 431
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
1291 .....GAGTCATACCTCAGCTATGAC...CAGCCCCGCA 1322
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
431 gGluTrpGlnArgAspAlaIleIleThrCysAsnProGluGluPheIleV 448
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
1323 G.....AAAAAGAGAAGGTTGTGAAAACTTCGGGTACAGCAGCTTGGAG 1366
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
448 alGluAlaLeuGlnLeuProAsnPheGlnGlnSerValGlnGluTyrArg 464
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
1367 AAAAGGACTTAAAGAAAGAGAC...TCTAAAGCACTAGTAAAAACTTG 1413
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
465 ArgSerAlaGlnAspGlyProAlaProAlaGluLysArgSerGlnTyrPr 481
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
1414 ACTCGGCTCAGAAATTACCACAGGCGCAATGAAACCAAGTCAGACAAAG.. 1461
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
481 oGluIlePheLeuGlyThrGlySerAlaIleProMetLysIleArgA 498
  
```

[illegible]

2229	CCACTCAAGTCTCTGAGGAGCTGGCCCTACGAAGGGCCCTAGTACCAGCAGTG	2278	
742LysValGlyValAlaPheAspHisMetIys	751	
2279	CCACATTAGCTCTGTCGCCAGCAGCTCTGTTTCTATGATCCAGGAAA	2328	
752	ValcysPheGlyAspPheThrMetProLysLeuIleProProLeuLy	768	
2329CCAGCTGTGAAGAAATTCGCCCG.....	2352	
768	sAlaLeuPheAlaGlyAspIleGluGluMetGluGluArgArgGluLySA	785	
2353ATGATGCCCAAGACATTAAAGCATTCGAAGACAGATTTCGCCGAC	2398	
785	rgGluLeuArgGlnValArgAlaAlaLeuLeuSerArgGluLeuAlaG1	801	
2399	GA.....TAAACAGAGCTTGCTTGAGGTTGAGTCTGGAAGC	2436	
801	yGlyLeuGluAspGlyGluProGlnGlnLysArgAlaHisThrGluGluP	818	
2437	AGGACTACAGGGACTATGGGGACGGGAGAGGATGTCCACAGAAGACC	2486	
818	ro 818		
2487	CG 2488		
seq_name: /cgn1_8/gcgdata/geneseq/geneseq/nal1988.DAT:N80456			
seq_documentation_block:			
ID	N80456 standard; DNA; 2987 bp.		
XX			
AC	N80456;		
XX			
DT	20-MAR-1991 (first entry)		
XX			
DE	Sequence of Rhodosporidium toruloides genomic DNA encoding		
DE	phenylalanine ammonia lyase (PAL).		
XX			
KW	phenylketonuria therapy; L-phenylalanine therapy; enzyme; ss		
XX			
OS	Rhodosporidium toruloides.		
XX			
Key	Location/Qualifiers		
FH	exon	354..595	
FT		/*tag= a	
FT		/label=EXON 1	
FT	intron	596..684	
FT		/*tag= b	
FT		/label=IVS 1	
FT	intron	772..785	
FT		/*tag= c	
FT		/label=IVS 2	
FT	intron	968..1024	
FT		/*tag= d	
FT		/label=IVS 3	
FT	intron	1602..1662	
FT		/*tag= e	
FT		/label=IVS 4	
FT	intron	1833..1891	
FT		/*tag= f	
FT		/label=IVS 5	
FT	intron	2181..2236	
FT		/*tag= g	
FT		/label=IVS 6	
FT	exon	2237..2821	
FT		/*tag= h	
FT		/label=EXON 7	
XX			
PN	W08802024-A.		
XX			
PD	24-MAR-1988.		
XX			
PF	08-SEP-1987; 87WO-GB00528.		

PI	Bradsher JN, Conway JW, Conway RC;
XX	
DR	WPI; 1997-192901/17.
DR	P-PSDB; W13849.
XX	
PT	DNA encoding RNA polymerase transcription factor, Elongin, 15, 18
PT	and 110 kDa subunits - used to modulate transcription rate of RNA
PT	polymerase
XX	
XX	Claim 28; Page 55-59; 99pp; English.
XX	
CC	cDNA clones (T59993-95) respectively code for elongin C, elongin B
CC	and elongin A (W13847-49), the 15 kDa, 18 kDa and 110 kDa subunits
CC	of rat Elongin, an RNA polymerase transcription factor which
CC	increases the rate of transcription elongation by RNA polymerase II.
CC	Elongin A cDNA was amplified from a rat brain cDNA library using
CC	primers T60005-06) based on tryptic peptides (W13959-60) of the
CC	isolated subunit. The cDNA clones are useful for the recombinant
CC	prodn. of Elongin. Human Elongin cDNA clones (see also T59996-98)
CC	have also been isolated.
XX	
XX	Sequence 3501 BP; 981 A; 861 C; 896 G; 762 T; 1 other;
XX	

173 TyrGluAspGluThrMetThrValTyrGln..... 182
 |||::|||::: |||||
 391 CAGGAGGAGGAAAGCGGAGGGAACATACCAGGAAAGCTGGCAAGCCTCAGG 440
 183IleProIleHisSerGluGlnArgArgGlyLysHisGlnProTrpG 198
 ||| |||::|||::: |||||
 441 CAGCCAGCCCTACAGCCCTGAGCACAGACAGAAAAGACAGAAAACCTTC 490
 198 InsSerProGluArgPro..... 203
 ::: |||||
 491 CTGAGCTTGAAGAGCCTCACAAAGTGCTCATGGTCACGACGAGGAGAT 540
 204LeuSerArgLeuSerProGluArgSerSerAspSerG 216
 ::::||| |||||
 541 GAAAGGAAGAGGTGTCAAAAGTGTCAACCATATTTCTTCAGACCCCGA 590
 216 uSerAsnGlu.....AsnGluProHisL 224
 |||::|||::: |||||
 591 GTCTGCTGACTATGTGCTCATGTTCAATCTCTCCACCTTCAAGTCCCATC 640
 224 euProHisGlyValSerGlnArgArgGlyValArgAspSerSerLeuVal 240
 ::: ||| |||||
 641 AAATGTATACAGACCTCTCTAGGTCCCGAGATGGCAGAACCCATC 590
 241 ValAlaPheIleCysLysLeuHisLeuLysArgGlyAsnPheLeuVal... 256
 |||:: ||| |||||
 691 GTCTCA.....CACCCGAAGCCTGGGAAAGTCCACAGTAA 725
 257LeuLysAlaLysGluMetGlyLeuProValGlyT 268
 |||:: ||| |||||
 726 TACCTTTTCAGGACAGACTAGGGGTAGTCACTGGGTGAGCACCAAGGA 775
 268 hrAlaAlaIleAla..... 272
 :::|||::|||
 776 AGGGGCTGTATGCCAAACAGCCACAAATCTTCCCAAGGAGAAA 825
 273 ...ProIleIleAlaValLysAspGlyLysSerIle..... 284
 |||:: ||| :::: |||||
 826 CGCCTGTGATGCCAGGGGATGAGNAGAGCTCTGTCATGGGCAGAGA 875
 285ThrHisGluGlyArgGluIleLeuAlaG 294
 ::|||:: |||:: |||:: |||:: |||:: |||:: |||:: |||
 876 GAAGTCACACAAAGCCTCTTCCAAAGAGGAGCGGACCGAGGCTACTCTCAG 925
 294 LuCluLeuCysThrProProAspProGlyAlaAlaPheValValGlu 310
 926 AGGACAGTGCACAGGAGAGAGTGCCTCCAGT.....GTTGTCAGAAA 969
 311 CysProAspGluSerPheIleGlnProIleCysGluAsnAlaThrPheG 327
 |||:: |||
 970 GAGAAGGACAGA..... 981
 327 nArgTyrGlnGlyLysAlaAspAlaProValAlaLeuValHisMetA 344
 ::|||:: |||:: |||:: |||:: |||:: |||:: |||
 982GAAGGCAAC.....AGCCTCAAGAAAGAAGTTCT 1009
 344 laProAla.....SerValLeuValAspSerArgTyrGlnGlnTrpMetGlu 359
 ::|||:: |||:: |||:: |||:: |||:: |||:: |||
 1010 CACCTGCCCTTAGATTGTGCTTCAGACACACCACTTTAAAGAGCCCAACAC 1059
 360 ArgPheGlyProAspThrGlnHisLeuValLeuAsnGluAsnCysAlaSe 376
 ::: |||||::: ::: |||||
 1060 AAG.....GACTCCGAGAAATCAAGTCTGCACAAAACACGACAG 1100
 376 rVal..... 377
 |||
 1101 TGTAGATAGCTGGACTCAGGACGAGGGACAGGAGACCCATTTACCAGAG 1150
 378HisAsnLeuArgSerHis.....LysIleGln 386
 ::|||:: |||:: |||:: |||:: |||:: |||:: |||
 1151 CCAAGGATAAAGTTCCCAACAACTGAAGGCTCAGAGGGGGAAGTAAG 1200

387 ThrGlnLeuAsnLeuIleHisProAspIlePheProLeuLeuThrSerph 403
 1201 ACTAACTCGGATCGAAAGTCAACAGGCTCACTCCCTCTAAAGTAGAAGAT 1250
 403 eaRgCys.....LysLysGluGlyProThrLeuSerValPrometValG 418
 1251 GCACATGGATGATGTTTGAGACGCCACCATGCTCTTT..... 1290
 418 lngGlyGluCysLeuLysTyrGlnLeuArgProArgGluUTrpGln 434
 1291GAGTCATACCTCAGCTATGAC...CAGCCCCGCAAG.....AAA 1326
 435 ArgAspAlaIleThrCysAsnProGluPheIleValGluAlaLe 451
 1327 AGAAGAAGGTGTGAAAACCTCCGGTCAGACACTGGAGAAAAGGACT 1376
 451 uGlnLeuProAsnPheGlnSerValGlnGluTyrArgArgSerAlaG 468
 1377 TAAAAGAAGGAC...TCTAAAGCACTAGTAAACACTTGAACCTGGCTC 1423
 468 lnaSPGlyProAlaProAlaGluLysArgSerGlnTyrProGluIle 484
 1424 AGAAATATTACCAAGCGAATGAAAACAAGTCAGACAAG..... 1461
 485 PheLeuGlyThrGlySerAlaIleProMetLysIleArgAsnValSerAl 501
 1462 ...TTGCAGCAGCTGGAGCCGACCCAGCAGGCTAGAAGGTCCCTAC 1508
 501 aThrLeuValAsnIleSerProAspThrSerLeuLeuLeuAspCysGlyG 518
 1509 TCATGTGTGTCAGCATTTGCCAGACATC..... 1536
 518 luClyThrPheGlyGlnCysArgHisTyrGlyAspGlnValAspArg 534
 1536 1536
 535 ValLeuGlyThrLeuAlaAlaValPheValSerHisLeuHisAlaAspHi 551
 1537CCCTGGCCGCCATA.....CAACCCAA 1559
 551 shiThrGlyLeuProSerIleLeuLeu..... 560
 1560 CTTACGTCCTCCCTCCCTCCCTCGATGTGATCTCCTCCTTCAGCCAAAGC 1609
 561GlnArgGluArgAlaLeuAlaSerLeuGly 570
 1610 GAAAGCTTTCTCTCACCCCGAGAGAGAGAGCTGGTTCACAGGA 1659
 571 LysProLeuHisProLeuValValAlaProAsnGlnLeuLysAlaTr 587
 1660 CGCAATGAATTTCTAAGATCAGGTGTATTCAGTTTCCAGTGTGCCTA 1709
 587 PheUlnGln.....TyrHisAsnGlnCysGlnGluValLeuHisH 601
 1710 TCTCCCAAAATGATGACCTTGGACCAAGCATGTATCCGGTGCTTAAGA 1759
 601 is.....IleSerMetIleProAlaLysCysLeu 610
 1760 ATAATATTGACTCCATCTTTGAAGTGGGAGGAGTCCCTATTCTGTCIT 1809
 611 GlnGluGlyAlaGluIleSerSerProAlaValGluArgLeuIleSerSe 627
 1810 GAA.....CCTGTCTTGAGAGGTGCACACCCGA 1838
 627 rLeuLeu.....ArgThrCysAsp.....LeuGluGluPheG 638
 1839 TCAGCTCTACCGAATAGAGGAATCCATCATGTATTAATTGAGGAACAG 1888
 638 lnaThrCysLeuValArgHisCysLysHisAlaPhe 649
 1889 ATCAGTTGTGAAAGTTCACTGTCCACGGGACTTT 1923

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
/cgnl_7/ptodata/1/ina/6A.COMB.seq:US-08-714-918-3 +	175.50	275.50	8.6e-08	2494	!
/cgnl_7/ptodata/1/ina/5A.COMB.seq:US-08-524-757-5 +	120.50	170.45	0.0609	3501	!
/cgnl_7/ptodata/1/ina/5A.COMB.seq:US-08-459-041A-1 +	120.00	155.87	0.3951	9759	!
/cgnl_7/ptodata/1/ina/6A.COMB.seq:US-08-999-733-1 +	120.00	155.87	0.3951	9759	!
/cgnl_7/ptodata/1/ina/6A.COMB.seq:US-09-036-987A-1 -	119.00	125.98	18.28	80161	!
/cgnl_7/ptodata/1/ina/5A.COMB.seq:US-08-247-901C-1 +	116.50	127.61	14.83	50341	!
/cgnl_7/ptodata/1/ina/5B.COMB.seq:US-09-073-904-1 +	116.50	127.61	14.83	50341	!
/cgnl_7/ptodata/1/ina/5A.COMB.seq:US-07-642-734C-3 -	115.00	137.02	4.44	20235	!
/cgnl_7/ptodata/1/ina/6A.COMB.seq:US-08-439-009A-3 +	115.00	137.02	4.44	20235	!
/cgnl_7/ptodata/1/ina/5A.COMB.seq:US-08-093-453B-1 +	114.50	145.83	1.43	9757	!
/cgnl_7/ptodata/1/ina/6A.COMB.seq:US-08-938-291A-1 +	114.00	160.16	0.2282	3109	!
/cgnl_7/ptodata/1/ina/PCTUS.COMB.seq:PCR-US95-16980-1 -	113.50	162.92	0.1600	2359	!
/cgnl_7/ptodata/1/ina/5B.COMB.seq:US-08-804-227C-7 +	113.50	123.81	24.15	44377	!
/cgnl_7/ptodata/1/ina/5B.COMB.seq:US-08-804-198-1 +	113.50	123.81	24.15	44377	!
/cgnl_7/ptodata/1/ina/6A.COMB.seq:US-09-320-878-21 +	113.00	149.63	0.8800	5970	!
/cgnl_7/ptodata/1/ina/PCTUS.COMB.seq:PCR-US93-03077-2 +	112.00	155.79	0.3994	3279	!
/cgnl_7/ptodata/1/ina/5A.COMB.seq:US-08-221-817-12 +	111.00	159.26	0.2560	2204	!
/cgnl_7/ptodata/1/ina/5A.COMB.seq:US-08-454-439-12 +	111.00	159.26	0.2560	2204	!
/cgnl_7/ptodata/1/ina/PCTUS.COMB.seq:PCR-US94-10487-1 -	111.00	159.26	0.2560	2204	!
/cgnl_7/ptodata/1/ina/5A.COMB.seq:US-08-471-117A-1 +	110.00	116.67	60.28	46999	!
/cgnl_7/ptodata/1/ina/5A.COMB.seq:US-08-042-747A-4 -	109.50	151.64	0.6799	3177	!
/cgnl_7/ptodata/1/ina/6A.COMB.seq:US-08-822-586-46 -	109.00	135.50	5.39	9960	!
/cgnl_7/ptodata/1/ina/PCTUS.COMB.seq:PCR-US95-04910-8 +	108.50	143.37	1.96	5153	!
/cgnl_7/ptodata/1/ina/5A.COMB.seq:US-08-363-300-1 +	108.00	145.70	1.46	4039	!
/cgnl_7/ptodata/1/ina/6A.COMB.seq:US-09-051-019-1 -	108.00	135.12	5.66	8931	!
/cgnl_7/ptodata/1/ina/6A.COMB.seq:US-09-320-878-19 +	108.00	115.65	68.77	38506	!
/cgnl_7/ptodata/1/ina/6A.COMB.seq:US-09-028-934-33 +	107.00	138.61	3.62	5996	!
/cgnl_7/ptodata/1/ina/5A.COMB.seq:US-07-903-466-1 -	106.50	146.84	1.26	3018	!
/cgnl_7/ptodata/1/ina/PCTUS.COMB.seq:PCR-US93-05794-1 -	106.50	146.84	1.26	3018	!
/cgnl_7/ptodata/1/ina/5A.COMB.seq:US-08-412-431-2 +	106.00	147.27	1.19	2729	!
/cgnl_7/ptodata/1/ina/5A.COMB.seq:US-08-623-679-2 +	106.00	147.27	1.19	2729	!
/cgnl_7/ptodata/1/ina/6A.COMB.seq:US-08-933-774-2 +	106.00	147.27	1.19	2729	!
/cgnl_7/ptodata/1/ina/5B.COMB.seq:US-08-804-227C-1 +	106.00	110.43	134.21	43280	!
/cgnl_7/ptodata/1/ina/6A.COMB.seq:US-08-716-351A-1 +	105.50	131.16	9.40	8535	!
/cgnl_7/ptodata/1/ina/5B.COMB.seq:US-08-452-075-1 +	104.50	143.99	1.81	2841	!
/cgnl_7/ptodata/1/ina/5A.COMB.seq:US-08-042-747A-7 -	104.50	143.52	1.93	2943	!
/cgnl_7/ptodata/1/ina/6A.COMB.seq:US-08-785-420-1 -	104.50	121.48	32.53	15378	!
/cgnl_7/ptodata/1/ina/5A.COMB.seq:US-08-764-233A-1 +	104.50	105.94	238.96	49377	!
/cgnl_7/ptodata/1/ina/5A.COMB.seq:US-07-903-466-2 -	104.00	149.41	0.9054	1767	!
/cgnl_7/ptodata/1/ina/PCTUS.COMB.seq:PCR-US93-05794-2 -	104.00	149.41	0.9054	1767	!

```
2154 AAAGAGGGCCATATATGGAAGTTACATTTTGGAAACGAGTGCAGGTTT 2203
492 ePrometLysileArgAsnValSerAlaThrLeuValAlaSerPro. 508
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2204 GCCTACAAAGAGAGAAATACCAAGCAATCGCCTTAATTTAGAACCAT 2253
509 .....AspThrSerLeuLeuLeuAspCysGlyGluGlyThrPheGlyGln 523
    : : : : : : : : : : : : : : : : : : : : : : : : : :
2254 ATTCCAATTCCATATGCTTTTCGACGTTGGTGAAGGTACACAGCACC 2303
524 LeuCysArgHisTyrGlyAspGlnValAspArgValLeuGlyThrLeuAl 540
    : : : : : : : : : : : : : : : : : : : : : : : : : :
2304 ATTTCATCAT.....GCAATTAATTAGGAAAGTGCAC 2338
540 aalaValPheValSerHisLeuHisAlaAspHisThrGlyLeuProS 557
    : : : : : : : : : : : : : : : : : : : : : : : : : :
2339 ACATATATTTACTCATATGCGATCATATTTTGGTTGGCCAG 2388
557 erileLeuGlnArgGluArgAlaLeuAlaSerLeuGlyLysProLeu 573
    : : : : : : : : : : : : : : : : : : : : : : : : : :
2389 GATTACTTCTAGTCGT.....TCTTTTCAGGGCGGTGAACAG 2426
574 HisProLeuValValAlaProAsnGlnLeuLysAlaTyrLeuGlnG 590
    ||||| : : : : : : : : : : : : : : : : : : : : : : : :
2427 AAGCGCTTACATGTTGGACCAAGGAATTAAGCATATGTGGAAT 2476
590 nTyrHisAsn 593
    : : : :
2477 GTCATGAAT 2486
```

seq_name: /cgnl_7/ptodata/1/ina/5A_COMB.seq:us-08-524-757-5

```
seq_documentation_block:
: Sequence 5, Application US/08524757
: Patent No. 5792634
: GENERAL INFORMATION:
: APPLICANT: Conaway, Ronald C.
: APPLICANT: Conaway, Joan W.
: APPLICANT: Bradsher, John N.
: TITLE OF INVENTION: RNA Polymerase Transcription Factor
: NUMBER OF SEQUENCES: 37
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: RICHARDS, MEDLOCK & ANDREWS
: STREET: 1201 Elm Street, Suite 4500
: CITY: Dallas
: STATE: TX
: COUNTRY: US
: ZIP: 75270-2197
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/524,757
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/US94/13621
: FILING DATE: 29-NOV-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/160087
: FILING DATE: 30-NOV-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Harre, John A.
: REGISTRATION NUMBER: 37,345
: REFERENCE/DOCKET NUMBER: B35006CIPCIP
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (214) 939-4500
: TELEFAX: (214) 939-4600
: INFORMATION FOR SEQ ID NO: 5:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3501 base pairs
```

```
: TYPE: nucleic acid
: STRANDEDNESS: both
: TOPOLOGY: both
: MOLECULE TYPE: CDNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 82..2403
: FEATURE:
: NAME/KEY: misc difference
: LOCATION: replace(2990,"")
: OTHER INFORMATION: /note="This base can be either a,
: OTHER INFORMATION: c, t, g."
US-08-524-757-5

alignment_scores:
  Quality: 120.50      Length: 763
  Ratio: 0.378        Gaps: 38
  Percent Similarity: 41.809  Percent Identity: 19.528

alignment_block:
US-09-434-382-2 x US-08-524-757-5
Align seg 1/1 to: US-08-524-757-5 from: 1 to: 3501
```

```
13 GlyArgThrMetSerGlnGlyArgThrIleSerGlnAlaProAlaArgAr 29
    ||| : : : : : ||||| ||| : : : : : ||||| |||
14 GCGGAGTTGCGGCGCCGAGGACGCTACGCGACCCAGCTTCGCGCGAGGAG 63
    ||| : : : : : ||||| ||| : : : : : ||||| |||
29 gGluArg.....ProArgLysAspProLeuArgHisLeuArg.ThrArg 43
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
64 GCCGCGCAGTGACAGCGATGCGCGGAGTCCGCGCTCCAAGTTGTGA 113
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
44 GluLysArgGlyProSerGlyCysSerGlyGlyProAsnThrValTyrLe 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
114 GAAGCTGACGGCGCGCTGGCTGGCAACCGGACCCCAA..... 152
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
60 uGlnValValAlaAlaGlySerArgAspSerGlyAlaAlaLeuTyrValP 77
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
153 .....GAGCTATTGAATAT 168
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
77 heSerGluPheAsnArgTyrLeuPhe.....AsnCysGlyGlu 89
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
169 TTGAAGAACTTTCGCTTACCTATTACGTAGACATTTCTGTGGAGAC 218
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
90 .GlyValGlnArgLeuMetGlnGluHisLysLeuLysValAlaArgLeuA 106
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
219 TGGGTGGGAAACAGTG..... 237
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
106 spAsnIlePheLeuThrArgMetHisTrpSerAsnValGlyGlyLeuSer 122
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
238 ..AACAGCTTT.....CGGAACAT...GAGCAAGTGGGAACCTTGCC 276
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
123 GlyMetIleLeuThrLeuLysGluThrGlyLeuProLysCysValLeuSe 139
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
277 AGAGACCTGTTGGCCAGTGGGAAGAGCTGTTCCA..... 312
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
139 rGlyProGlnLeuGluLysTyrLeuGluAlaIleLysIlePheSerG 156
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
313 .....GTAGAACGAAACAAATGAGGCT.....GAGGATC 340
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
156 lyProLeuLysGlyIleGluLeuAlaValArgProHisSerAlaProGlu 172
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
341 AGGATTTTGAGAAGCAATTCGCCAAGCGTCCCGAGAGATGTTCCCCAG 390
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
173 TyrGluAspGluThrMetThrValTyrGln..... 182
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
391 CAGGAGGAGGAAGCGAGGGGAACCTACCAGGAAAGCTGGCAAGCCTCAGG 440
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
183 ....IleProIleHisSerGluGlnArgArgGlyLysHisGlnProTyrG 198
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
441 CAGCCAGCCCTACAGCCCTGAGCACAGACAGAAAAAGCACAGAAAACTTC 490
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```



```

198 InSerProGluArgPro..... 203
491 CTGAGCTTGAAGGCGCTCACAAGTGGGTCTAGTGCACGAGAGAGAT 540
204 .....LeuSerArgLeuSerProGluArgSerSerAspSerG1 216
541 GAAAGGAAGAGGTGTCACAAGGTGTCACCAACCATATTCTTCAGACCCCGA 590
216 uSerAsnGlu.....AsnGluProHisL 224
591 GTCGTCTGACTATGTCATGTTCAATCTCTCCACCTTCAAGTCCCATC 640
224 euProHisGlyValSerGlnArgArgGlyValArgAspSerSerLeuVal 240
641 AAATGATACAGACCTCTCTAGTCCCGAGAGATGGACCAAGCAACCATC 690
241 ValAlaPheIleCysLysLeuHisLeuLysArgGlyAsnPheLeuVal... 256
691 GTCCTCA.....CACCGRAGCCTGGGAAGTCCACAGTAA 725
257 .....LeuLysAlaLysGluMetGlyLeuProValGlyT 268
726 TACCTTTCCAGACAGACTAGGGGTAGTCACCTGGGTGACCAAGGA 775
268 hAlaAlaIleAla..... 272
776 AAGGGCTGTTAGCCAAACAGCCACAAATCTCCCAAGGAGAA 825
273 ...ProIleIleAlaAlaValLysAspGlyLysSerIle..... 284
826 CCGCTGTGGATGCCAGGGGATGAGAGAGCTCTGTCATGGGAGAGA 875
285 .....ThrHisGluGlyArgGluIleLeuAlaG 294
876 GAAGTCACAAAGCCTCTTCCAAAGAGAGAGCGGAGGCTACTCTCAG 925
294 luGluLeuCysThrProProAspProGlyAlaAlaPheValValGlu 310
926 AGGACAGTGCACAGGAGAGAGCTGCCCTCCAGT.....GTTCTCAAGAAA 969
311 CysProAspGluSerPheIleGlnProIleCysGluAsnAlaThrPheG1 327
970 GAGAGGACAGA..... 981
327 nArgTyrGlnGlyLysAlaAspAlaProValAlaLeuValHisMetA 344
982 .....GAAGGCAAC.....AGCCTCAAGAAGAAGTTGT 1009
344 laProAla...SerValLeuValAspSerArgTyrGlnGlnTrpMetGlu 359
1010 CACCTGCCTTAGATGTTGCTTCAGACAAACCACTTTAAAAAGCCCAACAC 1059
360 ArgPheGlyProAspThrGlnHisLeuValLeuAsnGluAsnCysAlaSe 376
1060 ARG.....GACTCCGAGAAATCAAGTCTGACAAAAACACGAGAG 1100
376 rVal..... 377
1101 TGTATAGTGGTGGACTCAGGACGAGGACAGGAGACCACTTACCAGAG 1150
378 .....HisAsnLeuArgSerHis.....LysIleGln 386
1151 CCAAGGATAAAGTTCACCAACACCTGAAGGCTCAGAGGGGAAAGTAAGA 1200
387 ThrGlnLeuAsnLeuIleHisProAspIlePheProLeuLeuThrSerPh 403
1201 ACTAACTGGATCGAAGATCACAGGCTCACTCCCTTAAGTAGAAGAGAT 1250
403 eArgCys.....LysLysGluGlyProThrLeuSerValProMetValG 418
1251 GGACATGGATGATGAGTTGACCAAGCCCAACCATGTCCTTT..... 1290
418 lnglyGluCysLeuLeuLysTyrGlnLeuArgProArgArgGluTrpGln 434

```

```

1291 ....GAGTCATACCTCAGCTATGAC...CAGCCCGCAAG.....AAA 1326
435 ArgAspAlaIleIleThrCysAsnProGluGluPheIleValGluAlaLe 451
1327 AGAAGAAGGTTGTGAAAACCTCCGGTCAGACACTTGGAGAAAAAGGACT 1376
451 uGlnLeuProAsnPheGlnSerValGlnGluTyrArgArgSerAlaG 468
1377 TAAAAAGAAGGAC...TCTAAAAGCACTAGTAAAAACTTGAACCTCGCTC 1423
468 InAspGlyProAlaProAlaGluLysArgSerGlnTyrProGluIleIle 484
1424 AGAAATTACCCAAAGCGGAATGAACAAGTCAGACAAG..... 1461
485 PheLeuGlyThrGlySerAlaIlePrometLysIleArgAsnValSerAl 501
1462 ...TTGCACCGAGCTGGAGCCGACCCAGAGGCTTAGAAGGTCCCTAC 1508
501 aThrLeuValAsnIleSerProAspThrSerLeuLeuLeuAspCysGlyG 518
1509 TGATGTGCTGCCAGCATTCCAGACATC..... 1536
518 luGlyThrPheGlyGlnLeuCysArgHisTyrGlyAspGlnValAspArg 534
1536 ..... 1536
535 ValLeuGlyThrLeuAlaAlaValPheValSerHisLeuHisAlaAspHi 551
1537 .....CCCTTGCCCGCCATA.....CAAACCAA 1559
551 sHisThrGlyLeuProSerIleLeuLeu..... 560
1560 CTATCGTCCCTTCCCTCCCTCGAGTTGATCTCCTCCTTTTCACGCCAAGC 1609
561 .....GlnArgGluArgAlaLeuAlaSerLeuGly 570
1610 GAAAAGCTTCTCTTTCACCCCGAGGAAGAAGCTGGGTTCACAGGA 1659
571 LysProLeuHisProLeuLeuValValAlaProAsnGlnLeuLysAlaTr 587
1660 CCGAGNATGAATCTAAGATGCGAGGTGATTTCAGGTTCCAGTGCCTA 1709
587 pLeuGlnGln.....TyrHisAsnGlnCysGlnGluValLeuHisH 601
1710 TCTCCCAAAATGATGACCTTGACACCAAGTGTATCCGGGTGCTTAAGA 1759
601 is.....IleSerMetIleProAlaLysCysLeu 610
1760 ATAATATTGACTCCCATCTTTGAAGTGGGAGGATCCCTATTCTGTCTT 1809
611 GlnGluGlyAlaGluIleSerSerProAlaValGluArgLeuIleSerSe 627
1810 GAA.....CCTGTCTGGAGAGGTGCACACCCGA 1838
627 rLeuLeu.....ArgThrCysAsp.....LeuGluGluPheG 638
1839 TCAGCTCTACCGAATAGAGGAATGCAATCATGTATTAAATTGAGGAACAG 1888
638 InThrCysLeuValArgHisCysLysHisAlaPhe 649
1889 ATCAGTTGTGAAAGTTCACTGTCACCGGGGACTTT 1923

```

seq_name: /cgnl_7/ptodata/1/ina/5A_COMB.seq:US-08-459-041A-1

seq_documentation_block:

; Sequence 1, Application US/08459041A

; Patent No. 5663065

; GENERAL INFORMATION:

; APPLICANT: Frey, Teryl K.

; APPLICANT: Dominguez, Geraldina

; APPLICANT: Wang, Chin-Yen

; TITLE OF INVENTION: DNA Encoding Infectious Rubella Virus

```

:
: NUMBER OF SEQUENCES: 11
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Jones & Askew
: STREET: 191 Peachtree Street, 37th Floor
: CITY: Atlanta
: STATE: Georgia
: COUNTRY: USA
: ZIP: 30303
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/459,041A
: FILING DATE: 02-JUN-1995
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Greene, Jamie L.
: REGISTRATION NUMBER: 32,467
: REFERENCE/DOCKET NUMBER: 07362-0102
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (404) 818-3700
: TELEFAX: (404) 818-3799
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 9759 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: US-08-459-041A-1
:
: alignment_scores:
: Quality: 120.00 Length: 939
: Ratio: 0.323 Gaps: 48
: Percent Similarity: 39.617 Percent Identity: 19.276
:
: alignment_block:
: US-09-434-382-2 x US-08-459-041A-1
:
: Align seg 1/1 to: US-08-459-041A-1 from: 1 to: 9759
:
: 2 TrpAlaLeuCys.....SerLeuLeuArgSe 10
: ||| |||||
: 3943 TGGACCTGTGCCCCGACCGACGAGGGCTGGCCCGACGGCTACTACGACG 3992
:
: 10 rAlaAlaGly.ArgThrMetSerGlnGlyArgThrIleSerGlnAlaPro 26
: ||| ||| :|||:
: 3993 ACCTCGAGGTGCGCGCTCGGGGATGAGGCATGGCCCGCGCGCCCTC 4042
:
: 27 AlaArgArgGluArgProArgLysAspProLeuArgHisLeuArgThrAr 43
: ||| :|||:|||||
: 4043 GCATCACTCAACCGCCCTCGCAAA..... 4066
:
: 43 gGluLysArgGlyProSerGlyCysSerGlyGlyProAsnThrValTyrL 60
: ||||| :|||:
: 4067 .....GGCCCTTACATATACAGG 4085
:
: 60 euGlnValValAlaAlaGlySerArgAspSerGlyAlaAlaLeuTyrVal 76
: :|||:|||||
: 4086 TATGGAACATGGCGCAGGC.....GCTGGCAAGACTACCCGCATC 4126
:
: 77 PheSerGluPheAsnArgTyr...LeuPheAsnCysGlyGluGlyValGI 92
: :|||:|||||
: 4127 CTCGCTGCCCTTCAGCGCGAAGACCTTTACGTCTGC.....CCACCAA 4170
:
: 92 nArgLeuMetGlnGluHisLeuLysValAlaAlaArgLeuAspAsnIleP 109
: :|||:|||||
: 4171 TGGCCTCTCTGCACGAGATCCAGGCCAACTCCGCGCGCGCGAT..... 4213

```

```

109 heLeuThrArgMetHisTrpSerAsnValGlyGlyLeuSerGlyMetIle 125
4213 ..... 4213
126 LeuThrLeuLysGluThrGlyLeuProLysCysValLeuSerGlyProPr 142
:||||:|||||
4214 ATCGACATCAAGACGCGCCACCTACGAGCGCGGCTGACCAACCGCT 4263
142 oglnLeuGluLys.....TyrLeuGluAlaIleLysIlePheSerGly. 156
:||||:|||||
4264 CGCCGCCTACCGCCCATCTACATCGCATGAGCGGTCTACTCTCGCGCGG 4313
157 .....ProLeuLysGlyIleGluLeu 163
4314 AGTACTGCGCGTTCGTTGCCAGCCAAACACCGGAGGTGATCTGCGCTC 4363
164 AlaValArgProHisSerAlaProGluTyrGluAspGluThrMetThrVa 180
:||||:|||||
4364 GGTGATCGGGACAGTAGTGGCGCCACACTACGCCAATAACTGCCGACC.. 4411
180 lTyrGlnIleProIleHisSerGluGlnArgGlyLysHisGlnProT 197
:||||:|||||
4412 ....CCGTCCTCACCCTGGCTACCGAGCGCTCGGCCAC...ACTT 4454
197 rpGlnSerProGluArgProLeuSerArgLeuSerProGluArgSerSer 213
:||||:|||||
4455 GCGCTTCCCGACTGCTGGCGCGCCGCTCGCGCGGCTCGATTAT 4504
214 AspSerGluSerAsnGluAsnGluProHisLeuProHisGlyValSerGI 230
:||||:|||||
4505 GACATCGAG.....GGCGA 4518
230 nArgArgGlyVal.....ArgAspSerSerL 239
:||||:|||||
4519 GCGACCGGCACCTTCGCTCAACCTTTGGAGCGCGCGGCTCGACC 4568
239 euValValAlaPhe.....IleCysLysLeuHisLeuLysArg 251
|| :|||:
4569 TTCACCTCGCTTCGCGCGGAAACCGTGGCGCGCTTCAC...GAGGCT 4615
252 GlyAsnPheLeuValLeuLysAlaLysGluMetGlyLeuProValGlyTh 268
:||||:|||||
4616 GGCATACGCGCATACACCGTCCGCGAGGCCAGGTATGAGCGTCGGCAC 4665
268 rAlaAlaIle..... 271
:||||:|||||
4666 CGCTGCATCATCTAGCAGAGACGCGACGCTTGCCCTGCGCTGA 4715
272 .....AlaProIle 274
:||||:
4716 CACGCGACCTCGCCATCGTCAGCCTGACCGGGCTCCGACGCACTCTAC 4765
275 lIleAlaAlaValLysAspGly..... 281
:||||:|||||
4766 CTCACGAGCTCGAGCGGCTACTGCGCGCTCGGGGCTCAGCGGCTT 4815
282 .....LysSerIleThrHisGluGlyA 289
:||||:
4816 CCTCGACCGCGGCGCACTGGCGGAGCTCAAGGAGGTTCGCCCTGGCATTTG 4865
289 rgGluIleLeuAlaGluLeuCysThrProProAspProGlyAlaAla 305
:||||:|||||
4866 ACCGCGTGTGCGCTGAGCAGCAGCACCAACCGCTTGGCGCGCGCGCAC 4915
306 PheValValValGluCysProAspGluSerPheIleGlnProIleCysGI 322
:||||:|||||
4916 GGCATC.....CCGAGGCCCAAGACGTCGCGCGCTTCTGCC 4953
322 uAsnAlaThrPheGlnArgTyrGlnGlyLysAlaAspAlaProValAlaL 339
:||||:|||||
4954 CCGCACTCTGGAGGAGCTCTCTTCGCGCGGTGGCGCGCACCCC..... 4996

```

339 euValValHisMetAlaProAlaSerValLeuValAspSerArgTyrGln 355
||| |||
4997CATTACCGCGACCTCAACCGCGTGAACGCGGAGAA 5038
356 GlnTrpMetGluArgPheGlyProAspThrGlnHisLeuValLeuAsnGln 372
||| |||
5039GTGCGGTACATGCGCATCTCGGTCACCTG...CTCAACAA 5076
372 uAsnCysAlaSerValHisAsnLeuArgSerHisLysIleGlnThrGlnL 389
||| |||
5077 GAATCACACCGAGATGCCGGA.....ACGGAACGGCTTC 5111
389 euAsnLeuIleHisProAspIlePheProLeuLeuThrSerPheArgCys 405
||| |||
5112 TCAGTCCCGTTTCGCGG.....TGCGGCTACCGCGG 5143
406 LysLysGluGlyProThrLeuSerValProMetValGlnGlyGluCysLe 422
||| |||
5144 GCGAGGATGGTGCACCTCCGCACTGTGTGGCCCGCAGCACCGCG 5193
422 uLeuLysTyrGlnLeuArgProArg.....G 432
||| |||
5194 CCCTTTTCGCCACATCCACCGCGCGCTCACTGTGGGTGCGCCAGG 5243
432 LuTrpGlnArgAspAlaIleIleThrCysAsnProGluGluPheIleVal 448
||| |||
5244 AGTGGCGCATG.....ACGTACTTTCGGG 5266
449 GluAlaLeuGlnLeuProAsnPheGlnGlnSerValGlnGluTyrArgAr 465
||| |||
5267 GAACGGATCGACTCACTGATGTCACACGAGATGGGCGTGGCGCGG 5316
465 gSerAlaGlnAspGlyProAlaProAlaGluLysArgSerGlnTyrProG 482
||| |||
5317 GGAGCTCACCGACCGCTACCGG.....CGCGGCTATCTCTG 5351
482 LuIleIlePheLeuGlyThrGlySerAlaIleProMetLysIle...Arg 497
||| |||
5352 AG...ATCTTCGCGGCTGTGTACCGCGCAGAGCTGACGCTCCCGCGC 5398
498 AsnValSerAlaThrLeu.....ValAsnIleSerPro...As 509
||| |||
5399 TTCCTCAAGCCACCTTGAAGTGGTAGCGCGGCCCTCGGCCCGAGGA 5448
509 pThrSerLeuLeuLeuAspCys.....GlyGluGlyThrPheGlyGlnL 524
||| |||
5449 CACCGAG.....GACTGCACCGCGCTCAGGGGAAAGCGCGCTTG 5489
524 euCysArgHisTyrGlyAspGlnValAspArgValLeuGlyThrLeuAla 540
||| |||
5490 AGATCGGCGGTGGGCAAGGAGTGGTTTCAGGTTATG..... 5527
541 AlaValPheValSerHisLeuHisAlaAspHisHisThrGlyLeuProse 557
||| |||
5528TCCCGCATTTCCGCGCGATCCAGAA 5553
557 rIleLeuLeuGlnArgGluAlaLeuAlaSerLeuGlyLysProLeuH 574
||| |||
5554 GATCATCATG.....CGCGCTTCG 5573
574 isProLeuLeuValValAlaProAsn.....GlnLeuLysAla 586
||| |||
5574 GCCCGCAATTCCTGGGCGCTGGCCATACGAGCGCGGATGCGG 5623
587 TrpLeuGlnGlnTyrHisAsnGlnCysGlnGluValLeuHisIleLe 603
||| |||
5624 TGGTGCAG..... 5632
603 rMetIleProAlaLysCysLeuGlnGluGlyAlaGluIleSerSerProA 620
||| |||
5633GCCCATACACCAACG 5651
620 laValGluArgLeuIleSerSerLeuLeuArgThrCysAspLeuGluGlu 636

5652 CCATCGAG.....GTGACTTCACTAG 5674
637 Phe.....GlnThrCysLeuValArgHisCysLysHisAlaPheGln 650
||| |||
5675 TTCGACATGAACGACCCCTCGTACTCGGAGCGTGCAGCTCAGATTAG 5724
650 yCysAlaLeuValHisThrSerGlyTrpLysValValTyrSerGlyAspT 667
||| |||
5725 CGCGGCTCTCTTG.....G 5738
667 hrMetProCysGlu.....AlaLeuValArgMetGlyLysAspAla 680
||| |||
5739 GCGTCCCTTGGCGCGAAGACTACCGCGCGTCCGCGCGGACTACTGC 5788
681 ThrLeuLeuIleHisGluAlaThrLeuGluAspGlyLeuGluGluAl 697
||| |||
5789 ACCCTG.....CGGAACTGGGCTCCACTGAGACCGG 5820
697 aValGluLysThr.....HisSerThrTrs 706
||| |||
5821 CTCGAGCGCACAAAGCGGAGCGCGCGCTGCTGCACAAACACCAACC 5869
706 erClnAlaIleSerValGlyMetArgMet..... 715
||| |||
5870 ..GTGCCCATGTGCATGGCCATGCGCATGGTCCCAAGGCGTGCCTGG 5917
715 715
5918 GCGGGATTTCCAGGGTACGATATGTCATCTCTCCCCGAGGGCGC 5967
716AsnAlaGluPheIleMetLeuAsn. 723
||| |||
5968 GCGCAGCGCGGCACTCAAGTGGACCCCGCGAGGTGGGTCTTTGGCT 6017
724 ..HisPheSerGlnArgTyrAlaLysValProLeuPheSerProAsnPhe 739
||| |||
6018 TCACATCCCGGTGAAGCAGCGTGCAGCACCCCT.....ACCCCGAGCTTC 6061
740 SerGluLysValGlyValAlaPheAspHisMetLysValCysPheGlyAs 756
||| |||
6062 TCGGGCAGCTCGGCACCGCGCGCGG.....CTCTTCCATGA 6099
756 pPheProThrMetProLysLeuIleProProLeuLysAlaLeuPheAlaG 773
||| |||
6100 T.....GTCAACGACGCGCATCAAGGTCTTTGCGCGCGCTTCGACC 6143
773 LysPleGluGluMetGluArgArgGluLysArgGluLeuArgGln 789
||| |||
6144 CAGACGTGTTGAAGAACACAGAGTGGCCCTCTCTCGACCGCTCCGGGG 6193
790 ValArgAlaAlaLeu 794
||| |||
6194 GTCTACGCGGCTCTG 6208
seq_name: /cgn1/7ptodata/1/ina/6_COMB.seq:US-08-999-733-1
seq_documentation_block:
; Sequence 1, Application US/08999733
; Patent No. 6054573
; GENERAL INFORMATION:
; APPLICANT: Frey, Teryl K.
; APPLICANT: Pugachev, Konstantin V.
; APPLICANT: Abernathy, Emily S.
; TITLE OF INVENTION: Highly Infectious Rubella Virus Clones
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Jones & Askew, LLP
; STREET: 191 Peachtree Street, 37th Floor
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA

```

; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/999,733
; FILING DATE: 02-SEP-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/459,041
; FILING DATE: 02-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/093,453
; FILING DATE: 19-JUL-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/722,334
; FILING DATE: 28-JUN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Greene, Jamie L.
; REGISTRATION NUMBER: 32,467
; REFERENCE/DOCKET NUMBER: 07362-0103
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 818-3700
; TELEFAX: (404) 818-3799
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9759 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-999-733-1

alignment_scores:
  Quality: 120.00      Length: 939
  Ratio: 0.323        Gaps: 48
  Percent Similarity: 39.617      Percent Identity: 19.276

alignment_block:
  US-09-434-382-2 x US-08-999-733-1
  Align seg 1/1 to: US-08-999-733-1 from: 1 to: 9759

2 TrpAlaLeuCys.....SerLeuLeuArgSe 10
||| |||||
3943 TGGCACCTGTGCGCGACGACGAGGGCTGGCCCGCAGCGTACTACGAGC 3992

10 rAlaAlaGly ArgThrMetSerGlnGlyArgThrIleSerGlnAlaPro 26
||| |||
3993 ACCTGAGGTGCGCGCTCGGGATGACGCCATGCGCCGCGCGCCCTC 4042

27 AlaArgGluArgProArgLysAspProLeuArgHisLeuArgThrAr 43
|||
4043 GCATCAGTCCAAACGCCCTCGCAAA..... 4066

43 gLulysArgGlyProSerGlyCysSerGlyGlyProAsnThrValTyrL 60
|||||
4067 .....GGCCCTTACAATATCAGGG 4085

60 euGlnValValAlaAlaGlySerArgAspSerGlyAlaAlaLeuTyrVal 76
;
4086 TATGAACATGCGCGCAGGC.....GCTGGCAAGACTACCGGCATC 4126

77 PheSerGluPheAsnArgTyr...LeuPheAsnCysGlyGluGlyValGL 92
;
4127 CTCGCTGCCTTCACGCGCGGAAGACCTTTACGCTGTC.....CCACCAA 4170

92 nArgLeuMetGlnGluHisLysLeuLysValAlaArgLeuAspAsnIleP 109

```

```

;
4171 TGGGCTCTGTGACGAGATCCAGGGCAAACTCGCGCGCGCGAT..... 4213
;
109 heLeuThrArgMetHisTrpSerAsnValGlyGlyLeuSerGlyMetIle 125
4213 ..... 4213
;
126 LeuThrLeuLysGluThrGlyLeuProLysCysValLeuSerGlyProPr 142
;
4214 ATCGACATCAAGACGCGCCACCTACGAGCGCGCGCTGACGAAACCGCT 4263
;
142 oGlnLeuGluLys.....TyrLeuGluAlaIleLysIlePheSerGly. 156
;
4264 CGCGCCTACCGCGCATCTATCATGATGAGGCGCTTCTCTCGCGCGG 4313
;
157 .....ProLeuLysGlyIleGluLeu 163
;
4314 AGTACTGCGCGCTTGTTCGCCAGCAACACCGCGGAGGTGATGCGTC 4363
;
164 AlaValArgProHisSerAlaProGluTyrGluAspGluThrMetThrVa 180
;
4364 GGTGATCGGACAGTGGCGGCCACACTACGCCAATAACTGCCGACCC... 4411
;
180 lTyrGlnIleProIleHisSerGluGlnArgGlyLysHisGlnProT 197
;
4412 ....CCGCTCCTGACCGCTGGCCTACGAGCGCTCGCGCCAC...ACTT 4454
;
197 rpGlnSerProGluArgProLeuSerArgLeuSerProGluArgSerSer 213
;
4455 GCGCTTCCCGACTGTGGCGGCCCTCGCGCGCGGGCTCGATTAT 4504
;
214 AspSerGluSerAsnGluAsnGluProHisLeuProHisGlyValSerG 230
;
4505 GACATCGAG.....GGCGA 4518
;
230 nArgArgGlyVal.....ArgAspSerSerL 239
;
4519 CGCACCGGACCTTTCGCTGCAACCTTTGGACGCGCGCGAGTCTCGACC 4568
;
239 euValValAlaPhe.....IleCysLysLeuHisLeuLysArg 251
;
4569 TTCACCTCGCCTTTCGCGCGCAACCGTGGCGCGCTTCAC...GAGGCT 4615
;
252 GlyAsnPheLeuValLeuLysAlaLysGluMetGlyLeuProValGlyTh 268
;
4616 GCATACGCGCATACACCGTGGCGGAGGCCAGGTATGAGCGTGGCAC 4665
;
268 rAlaAlaIle..... 271
;
4666 CGCCTGATCATGTAGGACAGAGCGGACGCTTGCCTGCGCGTGA 4715
;
272 .....AlaProIle 274
;
4716 CACGGACCTCGCATCTGTCAGCTGACCGCGCGGCTCGGACGACTCTAC 4765
;
275 IleAlaAlaValLysAspGly..... 281
;
4766 CTCCACGAGCTCGAGGACGCTCCTGCGCGCTGCGGGGCTCAGCGCGTT 4815
;
282 .....LysSerIleThrHisGluGlyA 289
;
4816 CCTCGACGCGCGGACCTGCGGAGCTCAAGGAGGTTCCTCCGCTGCGATTG 4865
;
289 rgGluIleLeuAlaGluGluLeuCysThrProAspProGlyAlaAla 305
;
4866 ACCCGGTGTGCGCGTGGAGGAGGACCAACCGCTTCCCGCCCGCGCAC 4915
;
306 PheValValValGluCysProAspGluSerPheIleGlnProIleCysGL 322
;
4916 GGCATC.....CCGAGGCCCAAGACGTGCGCGCTTCTGCCC 4953
;
322 uAsnAlaThrPheGlnArgTyrGlnGlyLysAlaAspAlaProValAlaL 339
;

```



```

; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dow Agrosciences LLC Patent Department
; STREET: 9330 Zionsville Road
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: USA
; ZIP: 46268
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatenIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/036,987A
; FILING DATE: 09-MAR-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stuart, Donald R
; REGISTRATION NUMBER: 28,479
; REFERENCE/DOCKET NUMBER: 50,608
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317)337-4816
; TELEFAX: (317)337-4847
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 80161 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-09-036-987A-1

alignment_scores:
  Quality: 119.00      Length: 749
  Ratio: 0.378        Gaps: 50
  Percent Similarity: 42.056  Percent Identity: 21.762

alignment_block:
  US-09-434-382-2 x US-09-036-987A-1/rev ..

Align seg 1/1 to reverse of: US-09-036-987A-1 from: 1 to: 80161

      8 LeuArgSer.....AlaAlaGlyArgThrMetSerGlnGlyAr 20
      |||||
      57302 CTCGAAGCGGACAGGATCAGCGACTTGGCGGTCAGTAGTCGGCGCG 57253

      20 gThr...IleSerGlnAlaProAlaArgGluArgProArgLysAsp. 35
      ||||
      57252 CACATCAGTCGGGTGTCTCCAGCGCACCGGTGCGCGGAGAGCATCGG 57203

      36 .....ProLeu.....ArgHisLeuArgThrArg 43
      |||||
      57202 CACCACCGGTTACCGAACTCGACCGCCTGGCGGTTTCCCGGTACTCGT 57153

      44 GluLysArgGlyProSerGlyCysSer.....GlyGlyProAsnTh 57
      :|||
      57152 CCAGAATCGGTTCCATCCGTTGCGAGTGGAAGCATGGGAGACCGCGAAC 57103

      57 rValTyrLeuGlnValValAlaAlaGlySer.ArgAspSerGlyAlaAla 73
      :|||
      57102 CATTTGG.....TCCTAAGCCCTCGGC..... 57080

      74 LeuTyrValPheSerGluPheAsnArgTyrLeuPheAsnCysGlyGluGl 90
      :|||
      57079 ATCGAATTCGGCTCGGAGCTCCAGGACGACATCGCGGTGCGCGGAGAG... 57032

      90 yValGlnArgLeuMetGlnGlnHisLysLeuLysValAlaArgLeuAspA 107
      :|||
      57031 .....CACTACCGACTCGGCGCGGTGATC 57007

      107 sn.IlePheLeuThrArgMetHisTrpSerAsnValGlyGlyLeuSerGl 123

```

```

57006 GCAGCGATGCTGACCCGATCGCACACCCCGGAAAGCAGCGGTCCGACCGC 56957
123 yMetIleLeuThrLeuLysGluThrGlyLeuProLysCysValLeuSerG 140
56956 GCGCTCGTGGTGGCGGCGGAGCAT.....TGGCGCGCGGTG 56917
140 lyProProGlnLeuGluLysTyrLeuGluAlaIleLysIlePheSerGly 156
56916 GG.....CAATGCCTGCATCAGGTG.....GGC 56894
157 proLeuLysGlyIleGluLeuAlaValArgProHisSerAlaProGluTyr 173
56893 CCGAGCGCAACCAACCGCTGCGCGTCCGCAAGGACACACACCGACCG 56844
173 rGluAspGluThrMetThrValTyrGlnIleProIleHisSerGluGlnA 190
56843 CGTGGCGCGCAGCCCACTCGCGGACCGAGTGCC.....CAGC 56806
190 rArgGlyLys.....HisGlnProTyrGlnSerProGluArgPro 203
56805 ACCGCGCGCGCGCGCACGCCCGCAGCACCTCAGCAATCCAGGAGCGCGAC 56756
204 LeuSer.....ArgLeuSerProGluArgSerSerAspSerGl 216
56755 TTGAACCGCGAAGATTCGCCGCTGCGCCACAGCGTCCGCTCGAGCATCT 56706
216 uSerAsnGluAsnGluProHisLeuProHis..... 226
56705 GCTCATCAGCACCGCAACAGCACACCCCGCAAGCTCGCGCCCGCAGGTGCG 56656
227 .....GlyValSerGlnArgArgGlyVal 234
56655 CTTAGCGCGTCGAGGCTTCTCTCGAAGCGCTCGCGAAGCGCGGGAAGC 56606
235 ArgAspSerSerLeu.....ValValAlaPheIleCysLys 246
56605 CGCGCACAAACGCTTTCCTCCATCCCGCAGCCACTGACTGCTGACCGGAA 56556
246 s.....LeuHisLeuLysArgGlyAsnPheLeuValL 257
56555 AACGAATCCGATCGGCGCACCGAATTCGAGAGCCCGCTTATCACCCCC 56506
257 euLysAlaLysGluMetGly.....Leu 264
56505 GCGCGCTCGCGCATCGCGAGCGCTGCCAGCTGGACAGCAGTGCCTC 56456
265 ProValGlyThrAlaAlaIleAlaProIleIleAlaValLysAspGl 281
56455 CCGATCGGCGACCCCA.....CAGCAGG 56433
281 yLysSerIleThrHis.....GluGlyArgGlu.....IleL 292
56432 CCGCTTCGTCAGCACAGACGAGTGTGCGGAGTGAATACCCCAATCC 56383
292 euAlaGluGluLeuCysThrProAspProGly.....AlaAla 305
56382 AGCGGAAGCGCGCAGGACCGCGTCCAGCGCGGCTGCAGCGAGCGCG 56333
306 PheValValValGluCysProAspGluSerPheIleGlnProIleCysGl 322
56332 TTGGCGCTGCGCGCTGCGAG.....ACTCCTAGCCGACA 56298
322 uAsnAlaThrPheGlnArgTyrGlnGlyLysAlaAspAlaProValAlaL 339
56297 ACATCCACGGGATCACCGCTGGATG...CGCAGTACCGCGCGATCCGA. 56252
339 euValValHisMetAlaProAlaSerValLeuValAspSerArgTyrGln 355
56251 .....TTGGTCGATCTCCGC..... 56237
356 GlnTrpMetGluArgPheGlyProAspThrGlnHisLeuValLeuAsnGl 372
      |||

```



```

; JOURNAL:
; VOLUME:
; PAGES:
; DATE:
; DOCUMENT NUMBER:
; FILING DATE:
; PUBLICATION DATE:
; RELEVANT RESIDUES IN SEQ ID NO:
; US-08-247-901C-1

alignment_scores:
  Quality: 116.50      Length: 657
  Ratio: 0.407         Gaps: 34
  Percent Similarity: 43.531  Percent Identity: 21.461

alignment_block:
US-09-434-382-2 x US-08-247-901C-1

Align seg 1/1 to: US-08-247-901C-1 from: 1 to: 50341

2 TrpAlaLeuCySerLeuLeuArgSerAlaAlaGlyArgThrMetSerG1 18
|||||
|||||
|||||
2051 TGGACGGCTTCATCGAGCGCAGATCCACGGCGG.....TCCTG 2091

18 nGlyArgThrIleSerGlnAlaProAlaArgArgGluArgProArgLysA 35
|||||
|||||
|||||
2092 GTCGAGGACATCAAGAGTGTGTTACGCAGCGCGCGCGCGCGATCTT 2141

35 spProLeuArg.....HisLeuArgThrArg..... 43
|||||
|||||
2142 CACCGATAAATGGACGAACTTGGAAATCCCTGGGAGGTGCAGTAATGGC 2191

44 .....GluLysArgGlyProSerGlyCysSerGlyG1 54
|||||
|||||
2192 GCAGATCGAGCGGCACACACAAATCGAGGGTTCTCTGGCTGTCAGGTGG 2241

54 yProAsnThrValTyrLeuGlnValAlaAlaGlySerArg..AspSer 70
|||||
|||||
2242 CCCCTCGGGCGTT.....CGTCGACAGAAAC 2267

71 GlyAlaAlaLeuTyrValPheSerGluPheAsnArgTyrLeuPheAsnC 87
|||||
|||||
2268 GGC.....CACGTACTACCGCGGTCTCGGCCACGAAGTGGGCGG 2308

87 sGlyGluGlyValGlnArgLeu.....M 95
|||||
|||||
2309 TGGCGAGGGTCTCGAGATCCTCACTACGAGGGTCCAGGACCGTCGAGG 2358

95 etGlnGluHisLysLeu.....LysVal..AlaArgLeuAspAs 107
|||||
|||||
2359 TCTCCGAGGAGAGTCCGCCAAGCCACAGCGGCGCAGGAGTCCAGGCT 2408

107 nIlePheLeuThrArgMetHisTrpSerAsnValClyGlyLeuSerGlyM 124
|||||
|||||
2409 GAACCTTCGCCGAGGTCCGCAAGAGTGTGAGTGGCGCGGTCTCAGGCGG 2458

124 etIleLeuThrLeuLysGluThrGly.....LeuProLys..... 135
|||||
|||||
2459 CGACAGGAACCTACACAGAGGAGTGGAGTGAATTACCGGCTCCCGGTCT 2508

136 .....CysValLeuSerGlyProGlnLeuGluLysTyrLeuGluAl 150
|||||
|||||
2509 TTTCTGCTGCCAACTGGCTTTGCCAGATCAACGGTCCCGG..... 2547

150 alIeLysIlePheSerGlyProLeuLysGlyIleGluLeuAlaValArgP 167
|||||
|||||
2548 .....ATGGGTAAAGGCGCGCAACCGATGTCGAC 2575

167 roHisSerAlaProGluTyrGluAspGluThrMetThrValTyrGlnIle 183
|||||
|||||
2576 CACATCAGCGCGGAGGACGACCGCGGTCCCAATCTGCGAGGCGGCTG 2625

```

```

184 ProIleHisSerGluGln.....ArgArgGlyLysHis..... 194
|||||
|||||
2626 CCATGCTGTACGGCAAGAAATCAGCCGCCGAGGCGGTAGCCCGACGGC 2675

195 .....GlnProTrpGlnS 199
|||||
|||||
2676 GGAACCTTAGAGCCCGGAGGAGGACGACCCGGAACGCCATCTCTGGCGT 2725

199 erProGluArgProLeuSerArgLeuSerProGluArgSer..... 213
|||||
|||||
2726 CGATAAGCGGGCCAGGTGCCCGCTCCACCCAGGAGTGAACAGTGGGCAC 2775

213 ..... 213
2776 GCGAGGCCCAATCGGAAACGAGATGAAGCGGGTTCGTCGGAACACC 2825

213 ..... 213
2826 CGGACAGTCCAACGACACGATCCAGATGCCCGGTCTGGTGACGATCCCC 2875

214 ..AspSerGluSerAsnGluAsnGluProHisLeuProHisGlyValSer 229
|||||
|||||
2876 GAGATGGCGGATCTAAGCCACGACGCGCGCGCACGACGCTCCTCAAGA 2925

230 GlnArgArg..GlyValArgAspSerSerLeuValValAlaPhe..... 243
|||||
|||||
2926 CATGTACGAGTCTGATCAAGCAGTCGCGCGCTGAAGTACTACGAGCCGA 2975

244 .....IleCysLysLeuHisLeuLysArgGlyAsnPheLeuVal 256
|||||
|||||
2976 CCGACTGGCAGATGCCCGCATCGCCCTCTACACTTAAACCCAGGAATC 3025

257 LeuLysAlaLysGluMetGlyLeuProValGlyThr.....AlaAl 270
|||||
|||||
3026 ATCCGAGCGAGAACACGCGCAAGCCCGTGGCGCGATGAAGTCACTGCG 3075

270 alIeAlaProIleLeuAlaValLysAspGlyLysSerIleThrHisG 287
|||||
|||||
3076 CATCAACCATGCTCTCCGCGCTG.....CTGCTGACCG 3110

287 luGly.....ArgGluIleLeuAlaGluLeuLeuCysThrProAsp 301
|||||
|||||
3111 AAGGTGACCGCGCGCGCTCGCATCTCGAAGTCGAACGAGCACCCGCTGAC 3160

302 Pro..GlyAlaAlaPheValValGluCysProAspGluSerPheIleG 318
|||||
|||||
3161 CCGACAGCGGGA.....AGTCTGTGACGT 3186

318 lnProIleCysGluAsnAlaThrPheGlnArgTyrGlnGlyLysAlaAsp 334
|||||
|||||
3187 CACCGACGTGCTCAAGCAGCGCTCGCCAAAGCGAGCGCGGGAGCTGAT 3236

335 AlaProValAlaLeuValValHisMetAlaProAlaSerValLeuValAS 351
|||||
|||||
3237 GGTCCC.....CCGAGGGGTTCCTAGAGCGG 3262

351 pSerArgTyrGlnGlnTrpMetGluArgPheGlyProAspThrGlnHisL 368
|||||
|||||
3263 CTGCGGCTACCGCGCTC.....CCCTCGGGGTAGACATCGAA..... 3302

368 euValLeuAsnGluAsnCysAlaSerValHisAsnLeuArgSerHisLys 384
|||||
|||||
3303 .....AGGAACCATCATG 3314

385 IleGlnThrGlnLeuAsnLeuIleHisProAspIlePheProLeuLeuTh 401
|||||
|||||
3315 GCCGACCTCGGC...AACCCACTCGACCTCGAGATGCTCTGCTGGTCAC 3361

401 r.....SerPheArgCysLysLysGluGlyProThrLeuSerValProM 416
|||||
|||||
3362 AGGCGGGGACTTCGCTGG.....ACCATCGATTACCGCT 3396

416 etValGlnGlyGluCysLeuLeuLysTyrGlnLeuArgProArgGlu 432

```



```
||||| ..... ||| |||||
3397 GGGTCGGGAGAGCTGCTCCTCGAAGT.....CGAGACCGCGCGCA 3439
433 TrpGlnArgAspAlaIleIleThrCysAsnProGluGluPheIleValG1 449
||||| ||| |||||
3440 ACACACGCGCTGCATCAGCTCTATGTCACCGGGCGAC..... 3478
449 uAlaLeuGlnLeuProAsnPheGlnGlnSerValGlnGluTyrArgArgS 466
||||| ..... |||||
3479 .....CGGAGGCACGTACACGCTGAACGTCAC 3506
466 erAlaGlnAspGlyProAlaProAlaGluLysArg_SerGlnTyrProG1 482
||||| ..... |||||
3507 GCACCAACACCGCGCCATCGACTACACGAGCTGCGAGAAATCGCA 3556
482 uIleIle.....PheLeuGlyThrG 489
: ||| ..... |||||
3557 GGGCTGGCAGCGGACATCAACAGCGCTCTGGACGCGCGTGGAGCGG 3606
489 lySerAlaIleProMetLysIleArgAsnValSerAlaThrLeuValAsn 505
||||| ..... |||||
3607 GAAACGCTGCTGTCATCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3656
506 IleSer.....ProAspThrSerLeuLeuLeuAspCysG1 517
: ||| ..... |||||
3657 TTCAACCTCAACGCCAGCAAGCGCTCACCAGCAGTGTGTCAC..... 3701
517 yGluGlyThrPheGlyGlnLeuCysArgHisTyrGlyAspGlnValAspA 534
||||| ..... |||||
3702 .....ACGATCAACAGCGCGGACGAGCTTCTCGACACGTCGACC 3744
534 rgValLeuGly 537
: ||| ..... |||||
3745 AACTACTGGG 3755
```

seq_name: /cgn1_7/ptodata/1/ina/5B_COMB.seq:us-09-075-904-1

seq_documentation_block:

; Sequence 1, Application US/09075904

; Patent No. 5994137

; GENERAL INFORMATION:

; APPLICANT: Jacobs, et al.

; TITLE OF INVENTION: L5 SHUTTLE PHASMIDS

; NUMBER OF SEQUENCES: 1

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Amster, Rothstein & Ebenstein

; STREET: 90 Park Avenue

; CITY: New York

; STATE: New York

; COUNTRY: U.S.A.

; ZIP: 10016

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 3.5 inch 1.44 Mb storage diskette

; COMPUTER: IBM PC Compatible

; OPERATING SYSTEM: MS-DOS

; SOFTWARE: Word Processor (ASCII)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/075,904

; FILING DATE: May 11, 1998

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/247,901

; FILING DATE: May 23, 1994

; ATTORNEY/AGENT INFORMATION:

; NAME: Bogosian, Elizabeth A

; REGISTRATION NUMBER: 39,911

; REFERENCE/DOCKET NUMBER: 96700/475

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212) 697-5995

; TELEFAX: (212) 286-0854 or 286-0082

; TELEX: TWX 710-581-4766

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 50341

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE:

; DESCRIPTION: L5 shuttle phasmid sequence

; HYPOTHETICAL: NO

; ANTI-SENSE:

; FRAGMENT TYPE:

; ORIGINAL SOURCE:

; ORGANISM: L5 mycobacteriophage

; STRAIN:

; INDIVIDUAL ISOLATE:

; DEVELOPMENTAL STAGE:

; HAPLOTYPE:

; TISSUE TYPE:

; CELL LINE:

; ORGANELLE:

; IMMEDIATE SOURCE:

; POSITION IN GENOME:

; CHROMOSOME/SEGMENT:

; FEATURE:

; NAME/KEY:

; LOCATION:

; IDENTIFICATION METHOD:

; OTHER INFORMATION:

; PUBLICATION INFORMATION: No. 5994137e

; AUTHORS:

; TITLE:

; JOURNAL:

; VOLUME:

; PAGES:

; DATE:

; DOCUMENT NUMBER:

; FILING DATE:

; PUBLICATION DATE:

; RELEVANT RESIDUES IN SEQ ID NO:

US-09-075-904-1

alignment_scores:

Quality: 116.50 Length: 657

Ratio: 0.407 Gaps: 34

Percent Similarity: 43.531 Percent Identity: 21.461

alignment_block:

US-09-434-382-2 x US-09-075-904-1 ..

Align seg 1/1 to: US-09-075-904-1 from: 1 to: 50341

2 TrpAlaLeuCyssSerLeuLeuArgSerAlaAlaGlyArgThrMetSerG1 18

||||| ||| ||||| ||||| ||||| ||||| |||||

2051 TGGACGGCTTCATCGAGCGCGAGATCCACGGCGCG.....TCCTG 2091

18 nGlyArgThrIleSerGlnAlaProAlaArgGluArgProArgLysA 35

||||| ||||| ||||| ||||| ||||| ||||| |||||

2092 GTCGAGGACATCAAGAAGGTGCTGTTCCACGACGCCGCCGCGGATCTT 2141

35 spProLeuArg.....HisLeuArgThrArg..... 43

||||| ||||| ||||| ||||| ||||| ||||| |||||

2142 CACGATAAACTGGACGAACTTGGAAATCCCGTGGAGGTGCAGTAATGGC 2191

44GluLysArgGlyProSerGlyCyssSerGlyG1 54

||||| ||||| ||||| ||||| ||||| ||||| |||||

2192 GCAGATCGAGCGGACACACAAATCGAGGGTTCCTGGCTGTCGAGGTGG 2241

54 yProAsnThrValTyrLeuGlnValValAlaAlaGlySerArg_AspSer 70

||||| ||||| ||||| ||||| ||||| ||||| |||||

2242 CCCCTCGGGCGTT.....CGTCGACAGAAC 2267

71 GlyAlaAlaLeuTyrValPheSerGluPheAsnArgTyrLeuPheAsnCy 87

||||| ||||| ||||| ||||| ||||| ||||| |||||

2268 GGC.....CACGTACTGACCCGGCTGTCGCCACGACGAAGTGGGGCG 2308
87 sGlyGluGlyValGlnArgLeu.....M 95
|||||.....
2309 TGGCAGAGGTCTCGAGATCTCAACTACGAGGTCCAGGACCGTCGAGG 2358
95 etGlnGluHisLysLeu.....LysVal AlaArgLeuAspAs 107
.....
2359 TCTCCGACAGAGACTCGCGGAAGCCAGCGGCCAGCGAGTGCAGGCT 2408
107 nLePheLeuThrArgMetHisTrpSerAsnValGlyGlyLeuSerGlyM 124
|||.....
2409 GAACTTCGCGCGAGTTCGCAAGGAGTGAAGTGGCGCGCTCAGGCGG 2458
124 etIleLeuThrLeuLysGluThrGly.....LeuProLys..... 135
|||.....
2459 CGACAGGAACCTACCAAGGAGCTGGGAGCTGAATTACCGGCTCCCGTCT 2508
136CysValLeuSerGlyProGlnLeuGluLysTyrLeuGluAl 150
|||.....
2509 TTCTGCTGCCAACTGGCTTTCGAGATCAACGGTCCCG..... 2547
150 atIleLysIlePheSerGlyProLeuLysGlyIleGluLeuAlaValArgp 167
.....
2548ATCGGTAAAGCGCCCAACCGCATGTCGAC 2575
167 roHisSerAlaProGluTyrGluAspGluThrMetThrValTyrGlnIle 183
|||.....
2576 CACATCAAGCGGGGAAACACACACCGCGCTGCAATCTCAGGACGCGT 2625
184 ProIleHisSerGluGln.....ArgArgGlyLysHis..... 194
|||.....
2626 CCATGTCTCAGCGCAAGAAATCAGCCCGCGAGGCGTAGGCCGACGGC 2675
195GlnProTrpGlnS 199
2676 GGGAACTTAGACCGGAGGAGCGACCCAGACCGCATCTGGGCT 2725
199 erProGluArgProLeuSerArgLeuSerProGluArgSerSer..... 213
|||.....
2726 CGATAAGCGGGCAGGTGCGCCCTCCACCCAGGAGTGAACAGTGGGCAC 2775
213 213
2776 GCGAGGCCAATCGGAAACGAGATGAAGAGCGGTTGCTCGGAACACCC 2825
213 213
2826 CGACAGTCCACACCGACAGATCCAGATCCCGGTCTGGTGAGATCCCC 2875
214 ..AspSerGluSerAsnGluAsnGluProHisLeuProHisGlyValSer 229
|||.....
2876 GAGATGGCGATCTAAGCCACACCGCGCGCAGCCAGCTCGTCAAGGA 2925
230 GlnArgArgGlyValArgAspSerSerLeuValValAlaPhe..... 243
|||.....
2926 CATGTACGAGTGCATACGAGTGGCGCGCTGAAGTACTACGAGCCGA 2975
244IleCysLysLeuHisLeuLysArgGlyAsnPheLeuVal 256
|||.....
2976 CGGACTGGCAGATGCCCGGACTCGCCCTCTACACACTTAACGAGGAATC 3025
257 LeuLysAlaLysGluMetGlyLeuProValGlyThr.....AlaAl 270
|||.....
3026 ATCGCAGCGGACGACACGCGGCGCGTGGCGCGGATGAAGTCACTGC 3075
270 atIleAlaProIleIleAlaAlaValLysAspGlyLysSerIleThrHisG 287
|||.....
3076 CATCAACGAGATGCTCTCCGCGCTG.....CTGCTGACCG 3110
287 luGly.....ArgGluIleLeuAlaGluLeuLysCysThrProProAsp 301
|||.....
3111 AAGGTGACCGCGCGCTCCGAGTCTCGAAGTTCGACGAGACCGCGCTGAC 3160

302 Pro.GlyAlaAlaPheValValGluCysProAspGluSerPheIleG 318
|||.....
3161 CCGACAGCGCGGA.....AGTCTGTGACGT 3186
318 lnProIleCysGluAsnAlaThrPheGlnArgTyrGlnGlyLysAlaAsp 334
|||.....
3187 GACCGAGGTGCTCAAGCAGCGCTCGCCAAAGCGAGCGCGGAGTGTAT 3236
335 AlaProValAlaLeuValHisMetAlaProAlaSerValLeuValAs 351
|||.....
3237 GTTCCC.....CCGAGGGGTTCTTAGAGCCG 3262
351 pSerArgTyrGlnGlnTrpMetGluArgPheGlyProAspThrGlnHisL 368
|||.....
3263 CTGCCGCTACCAAGCGCTC.....CCCTCGGGGTAGACATCGAA..... 3302
368 euValLeuAsnGluAsnCysAlaSerValHisAsnLeuArgSerHisLys 384
|||.....
3303AGGAACCATG 3314
385 IleGlnThrGlnLeuAsnLeuIleHisProAspIlePheProLeuLeuTh 401
|||.....
3315 GCGGACCTCGGC...AACCCACTCGACCTCGAGATCTCTGCTGTGTAC 3361
401 r.....SerPheArgCysLysGluGlyProThrLeuSerValProM 416
|||.....
3362 AGCGCGGACTTCCGCTGG.....ACCATCGATTACCGCT 3396
416 etValGlnGlyCysLeuLeuLysTyrGlnLeuArgProArgArgGlu 432
|||.....
3397 GGGGTCCGGGAGAGTGTTCCTCGAACT.....CGAGACCGCGCGCA 3439
433 TrpGlnArgAspAlaIleIleThrCysAsnProGluGluPheIleValG 449
|||.....
3440 ACAACAGCGCTGCTACGTCTATGTACCGGGGCGAC..... 3478
449 uAlaLeuGlnLeuProAsnPheGlnGlnSerValGlnGluTyrArgArgS 466
|||.....
3479CGGAGGCACGTACACGCTGAACGTCAAC 3506
466 erAlaGlnAspGlyProAlaProAlaGluLysArg.SerGlnTyrProG 482
|||.....
3507 GGCACCAACACCCCGCCATCGACTACACGACGTGTCTCGGAGAAATCCGA 3556
482 uIleIle.....PheLeuGlyThrG 489
|||.....
3557 GGGGTGGCAGCGACATCCAAAGACGCTCTGGACGCGCCGTCGGAGCCG 3606
489 lySerAlaIleProMetLysIleArgAsnValSerAlaThrLeuValAsn 505
|||.....
3607 GAAACGCTGCTCGTCATCCGGTCTCGCTGTCCCTGCGTGGACACTGAAC 3656
506 IleSer.....ProAspThrSerLeuLeuLeuAspCysG 517
|||.....
3657 TTCAACCTCAACGCCCAAGCGCGCTCAACGAGCAGTGTGGTCAAC..... 3701
517 yGluGlyThrPheGlyGlnLeuCysArgHisTyrGlyAspGlnValAspA 534
|||.....
3702ACGATCAACAGCGCGGACGACTCTCTCGACACGTTCTCGACC 3744
534 rgValLeuGly 537
|||.....
3745 AACTACTTGGG 3755

seq_name: /cgnl_7/ptodata/1/ina/5A_COMB.seq.us-07-642-734C-3

seq_documentation_block:

; Sequence 3, Application US/07642734C
; Patent No. 5824513
; GENERAL INFORMATION:
; APPLICANT: Katz, L
; APPLICANT: Donadio, S

APPLICANT: Mcalpine, J B
TITLE OF INVENTION: Recombinant DNA Method for Producing
Thrombolytic Enzymes
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Edward H. Gorman
STREET: Abbott Laboratories D377/AP6D-2 One Abbott
CITY: Abbott Park
STATE: IL
COUNTRY: US
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/642,734C
FILING DATE: 17-JAN-91
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Dancigers, Andreas M
REGISTRATION NUMBER: 32652
REFERENCE/DOCKET NUMBER: 4952.US.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 708-937-9396
TELEFAX: 708-938-2623
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 20235 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Saccharopolyspora erythraea
STRAIN: NRRL 238
FEATURE:
NAME/KEY: CDS
LOCATION: 19..10722
OTHER INFORMATION: /codon_start= 19
OTHER INFORMATION: /function= "gene eryA"
OTHER INFORMATION: /product= "eryA ORF2 encoding modules 3 & 4 for
OTHER INFORMATION: 6-deoxyerythronolide B"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 19..4470
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 97..1482
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoacyl ACP synthase of module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1693..2670
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyltransferase domain module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3406..3921
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoreductase domain of module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4171..4428
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 3"
FEATURE:

NAME/KEY: misc_feature
LOCATION: 4471..10722
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: module 4"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4471..5847
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoacylACPSynthase domain of module"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 6054..7026
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyltransferase domain of module 4"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 7165..9216
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: dehydratase and enoylreductase domains m"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 9433..9984
OTHER INFORMATION: /function= "approximate span
OTHER INFORMATION: beta-ketoreductase of module 4"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10225..10483
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 4"
FEATURE:
NAME/KEY: CDS
LOCATION: 10723..20235
OTHER INFORMATION: /codon_start= 10723
OTHER INFORMATION: /function= "gene -eryA"
OTHER INFORMATION: /product= "orf3 encoding modules 5 & 6
OTHER INFORMATION: 6-deoxyerythronolide B formatio"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10723..15165
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10831..12174
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoacylACPSynthase domain of modul"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 12379..13350
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyltransferase domain of module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 14062..14610
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoreductase of module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 14857..15114
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 15166..20235
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: module 6"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 15172..16569
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoacylACPSynthase domain of modul"
FEATURE:
NAME/KEY: misc_feature

```

: LOCATION: 16768..17721
: OTHER INFORMATION: /function= "approximate span of
: OTHER INFORMATION: acyltransferase domain of module 6"
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 18379..18921
: OTHER INFORMATION: /function= "approximate span of
: OTHER INFORMATION: beta-ketoreductase domain of module 6"
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 19149..19398
: OTHER INFORMATION: /function= "approximate span of
: OTHER INFORMATION: acyl carrier domain of module 6"
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 19492..20235
: OTHER INFORMATION: /function= "approximate span of
: OTHER INFORMATION: thioesterase domain of module 6"
: US-07-642-734C-3

```

alignment_scores:

```

    Quality: 115.00      Length: 751
    Ratio: 0.389         Gaps: 40
    Percent Similarity: 39.414   Percent Identity: 21.039

```

alignment_block:

```
US-09-434-382-2 x US-07-642-734C-3/rev ..
```

```
Align seg 1/1 to reverse of: US-07-642-734C-3 from: 1 to: 20235
```

```
10 SerAlaAlaGlyArgThrMetSerGlnGlyArgThrIleSerGlnAlaApr 26
```

```
9943 TCGCGGGAAGCGCGCGCTCTCCGGAGGCGAG.....CC 9909
```

```
26 oAlaArg.....ArgGluArgProArgLysAspProLeuArgHisL 40
```

```
9908 AGCGGTTGAGCGACTCTTCGCGCGCGGTACACGCCCTGCCCGCGGCC 9859
```

```
40 eu.....ArgThrArg...GluLysArgGlyProSerGly 50
```

```
9858 GGCTAGCACCGACGCGCGGACGAGACAGCAGCAGGAAGCTCAGACCGG 9809
```

```
51 CysSer.....GlyGly ProAsnThrValTyrLeu..... 60
```

```
9808 TGTTCGCGTCAGTCGTGCAAGTTCCACGCGCGCTCGACCTTGCGCGCG 9759
```

```
61 .....GlnValValAlaAlaGlySerArgAspSerGly 71
```

```
9758 AGCACCTGCTCCACCGCGCTCGTCGATGGAGGTGACCGCCGTGCGC 9709
```

```
72 AlaAlaLeuTyrValPheSerGluPheAsnArgTyrLeuPheAsnCysG1 88
```

```
9708 .....CAGCACACCGCGCGGTGACAGCACCGG 9680
```

```
88 yGluGlyValGlnArgLeuMetGlnHisLysLeuLysValAlaArgL 105
```

```
9679 TCAGGCGCGGGAGCGCCGTCACAGCGCGAAGCGCGTCGCG..... 9634
```

```
105 euAspAsnIlePheLeuThrArgMetHisTrpSerAsnValGlyGlyLeu 121
```

```
9633 .....GTGCGGGGTGTC 9622
```

```
122 Ser.....GlyMetIleLeuThrLeuLys..... 129
```

```
9621 GCAAGCGCGATCTCCGCGGACGCGGCCCAAGTCTCGATCTCCGCGCGCA 9572
```

```
130 .....GluThrGlyL 133
```

```
9571 GCTGTCGGAACCGCGCGTCCGCGCGCGCGCTGCTACACGAGCAGG 9522
```

```
133 euProLysCysValLeuSerGlyProGlnLeuGluLysTyrLeuGlu 149
```

```
.....
```

```

9521 TCGCCGACCCCGTCTCGGTGAGAGGTGGCGCGGAGCAGCGCGTCCGAG 9472
150 AlaIleLysIlePheSerGlyProLeuLysGlyIleGluLeuAlaValAr 166
9471 CGT.....GCCGTGCCACCCCGTGCATCAGCACCGTCCGTCGCG 9434
166 gProHisSerAlaProGluTyrGluAspGluThrMetThrValTyrGlnI 183
9433 GTCCACCGCGCGCGGTGGGTGCAGCACAGCTTCCGACGTG...CTG 9387
183 leProIleHisSerGluGlnArgArgGlyLys..... 193
9386 CCGCGCTCATGTCTGCACACCGCGCGCGGATCCAGTCCAGTCCACGC 9337
194 .....HisGlnProTrpGln.Ser.....ProGluArgProLeuSer. 205
9336 CGATACCGGAGCGGTGCAGCTCCCGCGCGCCAGCAGCGCGACCT 9287
206 .....ArgLeuSerProGluArgSerSerAspSerGluSerAsnGluAsn 220
9286 CGCGCAGGATCTCCCGGAGCGGTGTACCCCGCTCGCGGAGTCCGAC 9237
221 GluProHisLeuProHis...GlyValSerGlnArgArgGlyVal..... 234
9236 GGGCGGTATCGGCGCGCGAAGTCCCGCGGTCCCGAGTCCGCTGTGCC 9187
235 .....ArgAspS 237
9186 CATCTCGAGAGACGCGCTCTCCGCGAGCAGATCCCGGAGTCTGCGA 9137
237 erSerLeuValValAlaPheIleCys...LysLeuHisLeuLysArgGlyA 253
9136 GCAGTTCGCGGTGAGGTGAGCAGCAGCTCCACCGCGCGGCTCGC 9087
253 snPheLeuValLeuLysAlaLysGluMetGlyLeuProValGlyThrAla 269
9086 GTGCGCTCCCGAACTTCGGCGGAACCGGTCTCCCGGAGGAGCGAT 9037
270 AlaIleAla.....ProIleIleAlaAlaVa 278
9036 GTGCTCGTCTCGAGACCGCGCGCGAGCTCCCGTGTGGCGGCGC 8987
278 lLysAspGlyLysSerIle.....ThrHisGluGlyA 289
8986 CGCGGTGGCAACACCTCCGCGCGCGTCCGCGCGCGCGCGCGCGCC 8937
289 rgGluIleLeuAlaGluLeuCysThrProProAspProGlyAlaAla 305
8936 GCCAT.....CGCGACACCGCTGC..... 8917
306 PheValValValGluCysProAspGluSerPheIleGlnProIleCysG1 322
8916 .....CGCTGCGTGTGAGCAGCAGCTGACCCGCGCGCGCGCC 8879
322 uAsnAlaThrPheGlnArgTyr.....GlnG 331
8878 CCGCGAGATCGTCAGCGCGTAATCGCGCGGTGTAGCGATGGCGACG 8829
331 lLysAlaAspAlaProValAlaLeuValHisMetAlaProAlaSer 347
8828 CCGCGCGGTGGGTGCTCCAGCGCGTCCGCGCTCGTGTGGAGGAGCGCG 8779
348 ValLeuValAsp...SerArgTyrGlnGlnTrpMetGluArg..... 360
8778 GTGATCGGTGACCGCGATCGCGCGAAGGCTCCCTGGAACAGCGCGACA 8729
360 ..... 360
8728 CCGGTCTCCGCGCGAAGCGGTCCAGTCCGCTCCGACCGCGGTGACG 8679
361 .....PheGlyProAspThrGln.....HisLeuVal 369
8678 ACACCGCGCGCTCGGTGCCCATGTCCGCTCTCGCGATACATCCCGAG 8629

```

```
370 LeuAsnGluAsnCysAlaSerValHisAsnLeuArgSerHisLysIleG1 386
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
8628 TCGAGAGAGACGTCGCGGAAGTTGACTCCGGTGGCGGCAC..... 8587
386 nThrGlnLeuAsnLeuIleHisProAspIlePhe.....ProLeuLeu 401
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
8586 .....GTCACCGCGACCTCGCGCGCGCGCGCGCGCGCGCGCGCG 8553
401 hrSerPheArgCysLysLysGluGluProThrLeuSerValProMetVal 417
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
8552 TCGAGTCGGGTGCGGCTCGAAGGCCACGAGTCGATCCGCCCGTCAC 8503
418 Gln.....GlyGluCysLeuLeuLysTyrGlnLeuArg...Pr 429
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
8502 CGCAGCAGAGAGTGGGTGCTCGCGAAGCGTCAGCGGAAGTGC 8453
429 oArgArgGluTrpGlnArgAspAlaIleIleThrCysAsnProGluGluP 446
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
8452 GCGCGAGAGCGCGCGCAGCAGACGTCGTCGCGCGCAGCAGCGAGCTGC 8403
446 helIleVal...GluAlaLeuGlnLeuProAsnPheGlnGlnSerValGln 461
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
8402 GGCTTGTCCGAACCGCAGCAGCATCCAGGTTCCGCGTCGTCGAG 8353
462 GluTyrArg.....ArgSerAlaGlnAspGlyPr 471
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
8352 GAGCAGCAGACCGTCCGCGGACTCGGCTCGCAGCAGCGATCACGCC 8303
471 o..... 471
8302 ACATCGCGCGCGCGCGGTTCGGGAACGCTGCTGCTCGGACAGGTG 8253
471 ..... 471
8252 ACCGCGCGGACGTGGGATGACGAGCGTCGCGCGCGGAGCTCTCCTG 8203
472 ...AlaProIleGluLysArgSerGlnTyrProGluIleIlePheLeuG1 487
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
8202 TTCGAGCGCGCGCGCAGCAGCGCGCGCGCGCGCGCGCGCGCG 8153
487 yThrGlySerAlaIleProMetLysIleArgAsnValSerAlaThrLeuV 504
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
8152 CC CGCGCGCG..... 8143
504 alaIleSerProAspThrSerLeuLeuLeuAspCysGlyGluGlyThr 520
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
8142 .....GGGTCGTCGCC 8131
521 PheGlyGlnLeuCysArgHisTyrGlyAspGlnValAspArgVal..... 535
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
8130 TTCGGG.....TTCGTAGCGGACAG...GACCGCGTCGGGTT 8096
536 .....LeuGlyThrLeuAlaAlaValPheValSerHisLeuHisAlaA 550
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
8095 CGGGTTCACGCGCGCGCAGCAGCGCGTCGAGGTCGAGGATCGCGCG 8046
550 sphHisThrGlyLeu.....ProSerIleLeuLeuGln 561
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
8045 ACCACAGCGCGTCGCGCGCGCGCTCGGCTAGCGCTCCCGTCAGAGC 7996
562 ArgGluArgAlaLeuAlaSerLeuGlyLysProLeuHisProLeuLeu 577
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
7995 GTGAGTCGCGCTCGCGCAGCGCGGTTCGGGTGCTGCCACTTC 7948
```

seq_name: /cgnl_7/ptdata/1/lna/6_COMB.seq:US-08-439-009A-3

seq_documentation_block:

; Sequence 3, Application US/08439009A
; Patent No. 6004787
; GENERAL INFORMATION:
; APPLICANT: Donadio, S
; APPLICANT: Katz, L

APPLICANT: Mcalpine, J B
TITLE OF INVENTION: Method of Directing Biosynthesis of
Specific Polyketides
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Steven F. Weinstock
STREET: Abbott Laboratories D377/AP6D-2 One Abbott
CITY: Park Rd
STATE: IL
COUNTRY: US
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/439,009A
FILING DATE: 11-MAY-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Casuto, Dianne
REGISTRATION NUMBER: 40,943
REFERENCE/DOCKET NUMBER: 4952.US.D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 847-938-3137
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 20235 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Saccharopolyspora erythraea
STRAIN: NRRL 238
FEATURE:
NAME/KEY: CDS
LOCATION: 19..10722 /codon_start= 19
OTHER INFORMATION: /function= "gene eryA"
OTHER INFORMATION: /product= "eryA ORE2 encoding modules 3 & 4 for
6-deoxyerythronolide B"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 19..4470
OTHER INFORMATION: /function= "approximate span of
beta-ketoacyl ACP synthase of module 3"
OTHER INFORMATION: module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 97..1482
OTHER INFORMATION: /function= "approximate span of
beta-ketoacyl ACP synthase of module 3"
OTHER INFORMATION: acyltransferase domain module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1693..2670
OTHER INFORMATION: /function= "approximate span of
acyltransferase domain module 3"
OTHER INFORMATION: module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3406..3921
OTHER INFORMATION: /function= "approximate span of
beta-ketoreductase domain of module 3"
OTHER INFORMATION: module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4171..4428
OTHER INFORMATION: /function= "approximate span of
acyl carrier domain of module 3"
OTHER INFORMATION: module 3"
FEATURE:
NAME/KEY: misc_feature

```

LOCATION: 4471..10722
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: module 4"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4471..5847
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoacylACPSynthase domain of module"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 6054..7026
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyltransferase domain of module 4"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 7165..9216
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: dehydratase and enoylreductase domains m"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 9433..9984
OTHER INFORMATION: /function= "approximate span
OTHER INFORMATION: beta-ketoreductase of module 4"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10225..10483
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 4"
FEATURE:
NAME/KEY: CDS
LOCATION: 10723..20235
OTHER INFORMATION: /codon_start= 10723
OTHER INFORMATION: /function= "gene =eryA"
OTHER INFORMATION: /product= "orf3 encoding modules 5 & 6"
OTHER INFORMATION: 6-deoxyerythronolide B formatio"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10723..15165
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10831..12174
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoacylACPSynthase domain of modul"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 12379..13350
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyltransferase domain of module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 14062..14610
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoreductase of module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 14857..15114
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 15166..20235
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: module 6"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 15172..16569
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoacylACPSynthase domain of modul"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 16768..17721

```

```

; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: acyltransferase domain of module 6"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 18379..18921
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: beta-ketoreductase domain of module 6"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 19149..19398
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: acyl carrier domain of module 6"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 19492..20235
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: thioesterase domain of module 6"
US-08-439-009A-3

alignment_scores:
    Quality: 115.00      Length: 751
    Ratio: 0.389        Gaps: 40
Percent Similarity: 39.414  Percent Identity: 21.039

alignment_block:
US-09-434-382-2 x US-08-439-009A-3/rev ..

Align seg 1/1 to reverse of: US-08-439-009A-3 from: 1 to: 20235

10 SerAlaAlaGlyArgThrMetSerGlnGlyArgThrIleSerGlnAlaPr 26
   ||| |||||::: ||| ||||::: ||| ||||::: ||| ||||::: |||
9943 TCGCGGAAGCGCGCGCTCTCCGGAGGCGAG.....CC 9909

26 oAlaArg.....ArgGluArgProArgLysAspProLeuArgHisL 40
   ||||| ||||:::||||||| ||||| ||||| ||||| |||||
9908 AGCGCGTTGAGGACTCGTTCGCGGCGCGGTACACGCCCTGCCCGGCC 9859

40 eu.....ArgThrArg...GluLysArgGlyProSerGly 50
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
9858 GCGTAGCACCGAGCGCGCGAGAACAGACAGACGAGAGAGTCTGACACCGG 9809

51 CysSer.....GlyGly.ProAsnThrValTyLeu..... 60
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
9808 TGTTCGCGTACGTCGTGTCAGGTTCCACGCCCGCGTCGACCTTGGCGCGC 9759

61 .....GlnValValAlaAlaGlySerArgAspSerGly 71
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
9758 AGCACCTGCTCCACCGCGCGTCTGTCGATGAGGTGACACGCCGCTCGCG 9709

72 AlaAlaLeuTyValPheSerGluPheAsnArgTyLeuPheAsnGly 88
   ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
9708 .....CAGCACACCGCGCGTGCACGACACCGCG 9680

88 yGluGlyValGlnArgLeuMetGlnHisLysLeuLysValAlaArgL 105
   ||::: ::::: ||||| ||||| ||||| ||||| ||||| |||||
9679 TCAGCGCGCGGGCAGCCCGTCCAGAGCGCCGAAAGCGCGTCCGCG.... 9634

105 euAspAsnIlePheLeuThrArgMetHisTrpSerAsnValGlyLeu 121
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
9633 .....CTCGCGCGGTGTC 9622

122 Ser.....GlyMetIleLeuThrLeuLys..... 129
   ::::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
9621 GCAAGCGCGATCTCGCGGAGACGCGCCCAAGTCTCTGATCTCGCGCGCA 9572

130 .....GluThrGlyL 133
   ::::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
9571 GCTCGTCGGAACCGCGCGTTCGCGCGCGCGCTGCTGACACGACGACGAG 9522

133 euProLysCysValLeuSerGlyProProGlnLeuGluTyrLeuGlu 149
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
9521 TCGCGCAGCCCGTGTCTGCGTGTGACAGAGTGGCGCGCAGCAGCGCTCCGAG 9472

```

150 AlalleLysIlePheSerGlyProLeuLysGlyIleGluLeuAlaValAr 166
 9471 CGTGCCGGTCCACCGGTGATCAGCACCGTCCGCTCCG 9434
 166 qProHisSerAlaProGluTyrGluAspGluThrMetThrValTyrGlnI 183
 9433 GGTCCACCGCGCGGGTGGTGCAGCAGAGTTCGCGAGGTG...CCTG 9387
 183 leProIleHisSerGluGlnArgArgGlyLys 193
 9386 CCCCGGTCATGTGTCGACAGCGCGCGCGGATCCCAAGTCCCAACGC 9337
 194HisGlnProTrpGln.Ser.....ProGluArgProLeuSer. 205
 9336 CGATACCGGGAGCGGTGAGCTCCCGCGCGCCAGCAGCGCCAGCACCT 9287
 206ArgLeuSerProGluArgSerSerAspSerGluSerAsnGluAsn 220
 9286 CGCGCAGGATCTCCCGAGCGGTGTACCCGCGCTCGCGAGGTGGAAC 9237
 221 GluProHisLeuProHis...GlyValSerGlnArgArgGlyVal..... 234
 9236 GGGCGGTATCGGCCCGGAGTCCCGCGGTCCCGCAGGTGCTGTGCC 9187
 235 237
 9186 CATCTCGACGAAGACGCGCTCTCCGCGAGCAGATCCGCGAGTCTGTGA 9137
 237 erSerLeuValValAlaPheIleCys..LysLeuHisLeuLysArgGlyA 253
 9136 GCAGTTCCTCCCGTGAGCGAGTTGACACCACTCCAGCCGCGCGCTCCG 9087
 253 snPheLeuValLeuLysAlaLysGluMetGlyLeuProValGlyThrAla 269
 9086 GTGCGCTCCCGAACTTCGCGCGAAACCGGTCTCCCGGAGGAAGCGAT 9037
 270 AlaIleAla.....ProIleIleAlaAlaVala 278
 9036 GTGCTGTGTGTGAGACGCGCGCGAGCTCCGCTGCTGTGCGCGGCG 8987
 278 llyAspGlyLysSerIle.....ThrHisGluGlyA 289
 8986 CGGCGGTGGCAACACTCCCGCCCGCTCGCGGGCCAGCGCGCGCC 8937
 289 rgGluIleLeuAlaGluGluLeuCysThrProProAspProGlyAlaAla 305
 8936 GCCAT.....GCCGACACCGCTGC..... 8917
 306 PheValValValGluCysProAspGluSerPheIleGlnProIleCysGl 322
 8916CGCTGGTGGATGAGACCGACTGACCCGCGCGCGAGCC 8879
 322 uAsnAlaThrPheGlnArgTyr.....GlnG 331
 8878 CGCGAGATCTGCAGCGCGTAATCGCGCTGGTGTAGGCGATGGGCACG 8829
 331 llyLysAlaAspAlaProValAlaLeuValValHisMetAlaProAlaSer 347
 8828 GCGCGGCGTGGGGTCTGCTCCAGCCGCTCGGCACTCGTGGAGGAGCGG 8779
 348 ValLeuValAsp...SerArgTyrGlnGlnTrpMetGluArg 360
 8778 GTGATCGGTGACCGGATCGCGCGGAAGGTTCCTGGAAACGCCGCGACA 8729
 360 360
 8728 CCCGCTCTCCCGCGGAAGCGCTCCAGTCCGCTCCGACCGCGGTGACG 8679
 361PheGlyProAspThrGln.....HisLeuVal 369
 8678 ACACCGCGGGCTCGGTCCGATGTCGCGCTTCTGCGGATACATCGCGAG 8629

```

370  LeuAnGluAsnCysAlaSerValHisAsnLeuArgSerHisLysIleG1 386
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
8628  TCGCAGGAGGAGCGTCGCGAAGTTGACTCCGTTGGCGGCAC..... 8587

386  nThrGlnLeuAsnLeuIleHisProAspilePhe.....ProLeuLeuT 401
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
8586  .....GTCCACCCGAGCCTCCGCCGCCCGCGGAGCGGCTGC 8553

401  hrSerPheArgCysLysGluGlyProThrLeuSerValProMetVal 417
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
8552  TCGACGTCGGGTGCGGGCTCGAAGCCAGCGAGTCGATCCGCCGTCACC 8503

418  Gln.....GlyGluCysLeuLeuLysTyrGlnLeuArg...Pr 429
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
8502  CGGCACGAGACGATTGGGTGCTTCCCGAAGCGTCACGCGGAAGGTGCGA 8453

429  oArgArgGluTTPGlnArgAspAlaIleIleThrCysAsnProGluGluP 446
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
8452  GCGGCGAGAGCGCGGACGACAGAGTCGTCCGCCGCAACCGGAGCTGC 8403

446  heIleVal...GluAlaLeuGlnLeuProAsnPheGlnGlnSerValGln 461
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
8402  GGGTTGTCGGAACCGCAGCGACGATCCCAAGTTCCGCGTCGCTGTCGAG 8353

462  GluTyrArg.....ArgSerAlaGlnAspGlyPr 471
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
8352  GAGCACGAACCGGTCGGCGACTCGCGCTGCCGACAGCGGATCACGCC 8303

471  o..... 471
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
8302  ACATCGCGCGCGCGGGTTCCGGNACCTGTCGTCGCGACACGGTG 8253

471  ..... 471
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
8252  ACCGCCCGGACGTGCGATGATCACCAGCTCGCGCGCGAGCTCCTCTG 8203

472  ...AlaProAlaGluLysArgSerGlnTyrProGluIleIlePheLeuG1 487
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
8202  TTCAGACCCAGCGCGCACAGAGCGCGCGCCAGAGGAGCGCGTCGCGG 8153

487  yThrGlySerAlaIleProMetLysIleArgAsnValSerAlaThrLeuV 504
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
8152  CCGCGGCGG..... 8143

504  aIAsnIleSerProAspThrSerLeuLeuAspCysGlyGluGlyThr 520
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
8142  .....GGGTCGTCGCC 8131

521  PheGlyGlnLeuCysArgHisTyrGlyAspGlnValAspArgVal..... 535
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
8130  TTCGG.....TTCGTAGCGGACCAAG...GACCGCGTCGGGT 8096

536  .....LeuGlyThrLeuAlaAlaValPheValSerHisLeuHisAlaA 550
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
8095  CGGGTTCACCGCACCCAGGCGGTGAGGTCCGAGGATCGGCGCGG 8046

550  sPHisHisThrGlyLeu.....ProSerIleLeuGln 561
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
8045  ACCACACGACCGGTCGAGCGCGCTGAGGTCCGAGGATCGGCGCGG 7996

562  ArgGluArgAlaLeuAlaSerLeuGlyLysProLeuHisProLeuLeu 577
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
7995  GTGCAGCTCGCCCTGCCACCGCGCGGTTCGGGCTGCTCCACTCTC 7948

seq_name: /cgnl_7/ptodata/1/ina/5A_COMB.seq:US-08-093-453B-1

```

```
seq_documentation_block:
; Sequence 1, Application US/08093453B
; Patent No. 5439814
; GENERAL INFORMATION:
; APPLICANT: Frey, Teryl K.
; APPLICANT: Domínguez, Geraldina
; APPLICANT: Wang, Chin-Yen
```



```
4913 GCAUC.....CCGAGGCCAAGAGCGCCGCCUUCUGGCC 4950
      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::
322 uAsnAlaThrPheGlnArgTyrGlnGlyLysAlaAspAlaProValAlaL 339
      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::
4951 CGCACUCUGGAGGAGCUCUCUCCGCGGCGGCCGCCACCC..... 4993
      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::
339 euValValHisMetAlaProAlaSerValLeuValAspSerArgTyrGln 355
      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::
4994 .....CAUACGCGGAGCCUCCACCGCGGUGAGUGAGGCGCAACGAGAA 5035
      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::
356 GlnTrpMetGluArgPheGlyProAspThrGlnHisLeuValLeuAsnG 372
      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::
5036 .....GUGGGUACAGGCGCAUCUCGCGUACCUG.....CUCAACAA 5073
      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::
372 uAsnCysAlaSerValHisAsnLeuArgSerHisLysIleGlnThrGln 389
      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::
5074 GAU.....CACACGAGAA 5087
      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::
389 euAsnLeuIleHisProAspIlePheProLeuLeuThrSer..... 402
      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::
5088 UG.....CCCGAAGGAGACGGGUUCACAGUGCGGUUGCGGC 5125
      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::
403 .....PheArgCysLysLysGluGlyProThrLeuSerValProMe 416
      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::
5126 GUGCGCGCUACGCGCGGAGGAGUGGUGACCCUCCGACUCUGU 5175
      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::
416 tValGlnGlyGluCysLeuLeuLysTyrGlnLeuArgProArgArg.... 431
      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::
5176 GGGCCCGCACGACCCCGCCUUCUCCGAGAGUCCACCCCGCGGUGCA 5225
      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::
432 .....GluTrpGlnArgAspAlaIleIleThrCysAsn 442
      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::
5226 CUGCUGGGUGCCCGAGAGUGGCGAUG..... 5254
      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::
443 ProGluGluPheIleValGluAlaLeuGlnLeuProAsnPheGlnInSe 459
      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::
5255 .....ACGUACUUGCGGAGAGGAGUCCACCUACUGUCUACACGCA 5298
      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::
459 rValGlnGluTyrArgSerAlaGlnAspGlyProAlaProAlaGluL 476
      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::
5299 GAUGGCGUGCGCGCGGAGGAGCACCACCGCUACGCG..... 5338
      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::
476 ysArgSerGlnTyrProGluIleIlePheLeuGlyThrGlySerAlaIle 492
      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::
5339 .....CGCGCUAUCCUGAG...AUCUUCGCGGCAUGUGUACCGGCCAG 5380
      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::
493 ProMetLysIle...ArgAsnValSerAlaThrLeu.....Va 504
      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::
5381 AGCCUGAGCGUCCCGCCUUCUCCAAAGCCACCUUGAAGUGGUGUACGCG 5430
      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::
504 lAsnIleSerPro...AspThrSerLeuLeuLeuAspCys.....GlyG 518
      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::
5431 CGCCUUGCGCGCCAGGACACCGAG.....GACUCCCGCGCGCUC 5471
      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::
518 luGlyThrPheGlyGlnLeuCysArgHisTyrGlyAspGlnValAspArg 534
      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::
5472 AGGGAAAGCGCGCUCUAGAUCCGCGGUGGCGCAAGGAGGUGGUGUACG 5521
      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::
535 valLeuGlyThrLeuAlaValPheValSerHisLeuHisAlaAspH 551
      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::
5522 GUUAUG.....UCCCGCA 5535
      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::
551 shiThrGlyLeuProSerIleLeuLeuGlnArgGluArgAlaLeuAlas 568
      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::
5536 UUUCCCGCGAUCCAGAGCAUCAUG..... 5563
      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::
568 erLeuGlyLysProLeuHisProLeuValValAlaProAsn..... 582
      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::
5564 .....CGCGCUUGCGCGCGCAAUCCUUGGCGCGCGCGCAUACG 5605
      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::
583 .....GlnLeuLysAlaTrpLeuGlnGlnTyrHisAsnGlnCysGlnG 597
      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::
```

```
5506 GAGCCCGAGGUGCAUCCGUGGUGGAG..... 5632
      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::
597 uValLeuHisHisIleSerMetIleProAlaLysCysLeuGlnGluGlyA 614
      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::
5633 ..... 5633
      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::
614 laGluIleSerSerProAlaValAlaGluArgLeuIleSerSerLeuLeuArg 630
      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::
5634 CCUAUACACACCAACGCAUCGAG..... 5659
      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::
631 ThrCysAspLeuGluGluPhe.....GlnThrCysLeuValArgH 644
      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::
5660 .....GUCACAUUCACUGAGUUGGACAUACCAACCCUCCUACUCGGA 5706
      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::
644 sCysLysHisAlaPheGlyCysAlaLeuValHisThrSerGlyTrpLys 661
      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::
5707 CGUCGAGCUGAGAUUAGCGCGCUCUCUG..... 5737
      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::
661 alValTyrSerGlyAspThrMetProCysGlu.....AlaLeuVal 674
      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::
5738 .....GGCCUCCUUGCGCGAAGACUACCGCGGCUC 5770
      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::
675 ArgMetGlyLysAspAlaThrLeuLeuIleHisGluAlaThrLeuGluAs 691
      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::
5771 CGCGCGGAGCUCUACUGACCCUG.....CGCGAACU 5802
      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::
691 pGlyLeuGluGluAlaValGluLysThr..... 701
      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::
5803 GGGCUCCACUGAGACCGGUGCGGACGACAGCGGCGGCGCCCGCCACGC 5852
      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::
702 .....HisSerThrThrSerGlnAlaIleSerValGlyMetArgMet... 715
      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::
5853 UGCUGCACAAACACACC...GUGCCAUUGUGCAUGGCGCAUGCGCAUGGUC 5899
      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::
715 ..... 715
      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::
5900 CCAAAAGCGGUGCGUGGCGCGGAAUUDCCAGGGUGACAUUGUCAU 5949
      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::
716 .....AsnAlaG 718
      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::
5950 CUUCCUCCGAGGCGCGGACGCGGACUACAAGUGACCCCGCGC 5999
      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::
718 luPheIleMetLeuAsn...HisPheSerGlnArgTyrAlaLysValPro 733
      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::
6000 AGGUGGCUUGUUGCUUCCACUCCGUGAAGCAGCUGAGCACCCCU 6049
      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::
734 LeuPheSerProAsnPheSerGluLysValGlyValAlaPheAspHisMe 750
      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::
6050 .....ACCCCGAGCUUUGCGGCGGACGCGGACCCCGCGCGC..... 6088
      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::
750 tLysValCysPheGlyAspPheProThrMetProLysLeuIleProL 767
      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::
6089 .....CUUCCUACUAGU.....GUCAGCACCGGCAUCAAGGUGC 6125
      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::
767 euLysAlaLeuPheAlaGlyAspIleGluMetGluGluArgArgGlu 783
      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::
6126 UUGGCGCGCUUUGCAGCCAGCAGCUGUAGAGAACAGCAGGUGGCCUC 6175
      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::
784 LysArgGluLeuArgGlnValArgAlaAlaLeu 794
      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::
6176 CUGCAGCCGCGGCGGCGGUCUACGCGGCUCUG 6208
      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::      ::::
```

seq_name: /cgml_7/ptodata/1/ina/6_COMB.seq:US-08-938-291A-1

seq_documentation_block:

; Sequence 1, Application US/08938291A

; Patent No. 6117673

; GENERAL INFORMATION:

; APPLICANT: Lev, Sima

; APPLICANT: Plovman, Gregory D.

; APPLICANT: Schlössinger, Joseph

; TITLE OF INVENTION: RDB PROTEINS AND RELATED

; TITLE OF INVENTION: PRODUCTS AND METHODS


```

1523  ACCTGCTCGCTACAGAGGTTCCACTGGCGGATGGCGAGTCCCTCT 1572
331  .....
1573  CTTGCTGATGCCCTACACACACAGCCCTCTTCTTGGAGGGCAGCT 1622
332  ysAlaAspAlaProValAlaLeuValValHisMetAlaProAlaSerVal 348
1623  CCGGACAGAGCCCGCCTCTG.....GATGCCCTCGCTCGGCC 1663
349  LeuValAspSerArgTyrGlnGlnTrpMetGluArgPheGlyProAspTh 365
1664  CTTACGGCTCGAGTTTCAGGCCCGCCAGCAGGATGAGGAGGGGAG 1713
365  rGlnHisLeuValLeuAsnCysAlaSerValHisAsnLeuArgS 382
1714  CTCCACAGCAGAGCTCGGATCTCGACAGCATGGCACCCGTTGGTG 1763
382  erHisLysIle.....GlnThrGlnLeuAsn 390
1764  CTTCCCGCATACAGCCAAAGTGGTGGGGAAGCAAGAGATCGACTATGCC 1813
391  LeuIleHisProAspIlePheProLeuLeuThrSerPheArgCysLysLy 407
1814  CTGTACTGCCCTGATGTC.....CTCACGGCCTTC..... 1843
407  sGluGlyProThrLeuSerValProMetValGlnGlyGluCysLeuLeuL 424
1844  .....CCACCGTGGCCCTGCC..... 1861
424  ySTyrGlnLeuArgProArgArgGluTrpGlnArgAspAlaIleLeuThr 440
1862  .....CACCTCTTCCAGCCGACGTTACTGGAGTCCACAGACGTGGTGCC 1906
441  CysAsnProGluGluPheIleValGluAlaLeuGlnLeuProAsnPheGl 457
1907  .....TTCATCTGTAGA.....CAGGTAATGCGCTATCA 1935
457  nGln...SerValGlnGluTyrArgArgSerAlaGlnAspGlyProAlaP 473
1936  GAGCGTGAACATCAAGAAAGCGCCCGCTGGACCTCGCAGCAGCTAGTC 1985
473  roAlaGluLysArgSerGlnTyrProGluIleIlePheLeuGlyThrGly 489
1986  CTGCCAACCCCGGAGAAGTGG.....CTTCGT 2014
490  SerAlaIleProMetLysIleArgAsnValSerAla 501
2015  AAGCGGACTCAGCTCAAGCTGAGGATGTCACGGCT 2050

```

seq_name: /cgnl_7/ptodata/1/ina/PTCUS_COMB.seq:PCT-US95-16980-1

```

seq_documentation_block:
; Sequence 1, Application PC/TUS9516980
; GENERAL INFORMATION:
; APPLICANT: Brigham & Women's Hospital
; TITLE OF INVENTION: CONTROLLING TRAF-MEDIATED SIGNALS
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street Suite 3100
; CITY: Boston
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/16980

```

```

; FILING DATE:
; PRIOR APPLICATION NUMBER: 08/367,540
; FILING DATE: 30-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Freeman, John W.
; REGISTRATION NUMBER: 29,066
; REFERENCE/DOCKET NUMBER: 05311/014W01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)542-5070
; TELEFAX: (617)542-8906
; TELEX: 100254
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2359 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 151..1854
; PCT-US95-16980-1

alignment_scores:
  Quality: 113.50      Length: 672
  Ratio: 0.425        Gaps: 38
  Percent Similarity: 39.732  Percent Identity: 21.726

alignment_block:
US-09-434-382-2 x PCT-US95-16980-1/rev ..

Align seg 1/1 to reverse of: PCT-US95-16980-1 from: 1 to: 2359

166 ArgProHisSerAlaProGluTyrGluAspGluThrMetThrValTyrGl 182
186 CGTCCGCTTCTGCC.....GCTTCTCCGCTCTCACA 1955
182 nIleProIleHisSerGluGlnArgArgGlyLysHisGlnProTrpGlns 199
1954 AGGTCTCTTTCTGAGCGCGAGGACCTCAGTGAAGACAGACAGCCGTTCAAT 1905
199 erProGlu.....ArgProLeuSer 205
1904 CCCCCAGAGGAGTTGCTTCTGCTGCTAAATCCACCTCCCGAGCTACTTATCA 1855
206 ArgLeuSerProGluArgSerSerAspSerGluSerAsnGluAsn..... 220
1854 GGGATCGGCAGATCCGAAAGTATCCACTATGACTTTAATAAAAAATTGTAT 1805
220 ..... 220
1804 CATCTTTAATATATGTCCTTCTAGAACAGTTTGGGCCCAAGACT 1755
221 .....GluProHisLeuProHisGlyValSerGlnArgArg 232
1754 GGGCAGCCAGAGCGGATATCTCTCTCCAGTGGGCTTCTTGAAGCTGCT 1705
233 GlyValArgAspSerSerLeuValValAlaPheIleCys..... 245
1704 GCTGTTGG.....GTCGGCTTGAATGCATCTCCCAAT 1670
246 .....LysLeuHisLeuLysArgGlyAsnPheLeuVal 257
1669 GAGCTCGAGAGGACCCCTGATCATCATCAGTATGAGTCTCTTCTGCTTA 1620
257 eulysAlaLysGluMetGlyLeu..... 264
1619 AACGGCCAGGAGCAGGCGCATCATATTCTCCACGCGATGATGACAAAAA 1570
265 .....ProVal..... 266

```

1569 CAGCGACAAGTCGGCTCCCGCTCCCGATCCCGCTCCCGCTCAGGTAGACCC 1520
267 .GlyThrAlaAlaIleAlaProIleIleAla.....AlaValL 279
||||| ||||| : : : : : |||||
1519 TGGCACACACTCTTATAGCAAGTAACAGCAGTGTAGAAAGGCTGGCTGTAA 1470
279 ysaspGlyLysSerIleThrHisGluGlyArgGluIleLeuAlaGluGlu 295
|| ||||| : : : : : |||||
1469 AG...GCACAGGGTCTCCCATACACGCG..... 1444
296 LeuCysThrProAspProGlyAlaAlaPheValValValGluCysPr 312
||| ||||| : : : : : |||||
1443 CTCCTGCTTCGCGG.....CTGTAGTCGGAATCTTCC 1409
312 oAspGlu...SerPheIleGlnProIleCysGluAsnAlaThrPheGlnA 328
||||| ||||| : : : : : |||||
1408 AGATGAGCACTCCATTCTAGTGGCGGTCTCCAGGACCTGGAAGCGCAGG 1359
328 rgTyrglnGlnLysAlaAspAlaProValAlaLeuValValHisMetAla 344
: : : ||| ||||| |||||
1358 TCCATGTGCGCTAGGCGGAT.....GTCCGTGCAC..... 1330
345 ProAlaSerValLeuValAspSerArgTyrglnGlnTrpMetGluArgPh 361
1330 1330
361 eGlyProAspThrGlnHisLeuVal.....LeuAsnG 372
||||| ||||| |||||
1329ACTCAGCACTGTGGTCAATGCGCGCTCAGCTGGGACTCCAGC 1290
372 luAsnCys.....AlaSerValHisAsnLeuArg 381
: : ||| ||| ||||| |||||
1289 AGGCCTGTGTCCGAGCCACTTGGCCCGCGCTCTTGTCCACGCTCTCCAG 1240
382 ..SerHisLysIleGlnThrGlnLeuAsn...LeuIleHisProAspII 396
: : ||| ||| ||||| : : : : : |||||
1239 CTCGGTCACGGGTCTGGAGGACTCCAGCTGCTCTTCATGCTGTCTG 1190
396 ePhe...ProLeuLeuThrSerPheArgCysLysLysGluGlyProThrL 412
||| ||||| : : : : : |||||
1189 CTCTCTCCAGTCTG.....CCGGAAGGCGCGATC 1158
412 euSerValProMetValGlnGlyGluCysLeuLeuLysTyrglnLeuArg 428
|| ||| : : : : : |||||
1157 TCCGTGCAAGCTCCTTCAGTTCCTGCTGTGGCTGTATCACTCGCTG 1108
429 ProArgArgGluTrp.....GlnArgAspAlaIleIleThrCy 441
||||| : : : : : |||||
1107 TAATGAAGATTTGGATTCATTATTTCGAAGCATTTCCCTTTGTCTCT 1058
441 sAsnProGluGluPheIleValGluAlaLeuGlnLeuProAsnPheGlnG 458
||| : : : : : ||| : : : : : |||||
1057 CAAT...TTCAATTCAAAGCTACATATCTGATTGTGCAAACTTTGTATG 1011
458 lnSerValGlnGluTyrglnArgSerAlaGlnAspGlyProAlaProAla 474
||| ||| ||| ||||| |||||
1010 CTCGTGTTTTTCTACACTTTCATTCTGCAACAAAGGAACCTTCTTTTC 961
475 GluLysArgSer.....GlnTyrgluIleIlePheLeuGlyTh 488
||||| : : : : : ||| : : : : : |||||
960 GAGGAGTTGCTCCACTCTTCAGCAGGTGACGTGTGACGCGGAGC 911
488 rGly.....SerAlaIleProMetLysIleArgAsnValSerA 501
||| : : : : : ||| : : : : : |||||
910 TGGCCTCGTGGCGCTTCACTGCTGTTGTTGTCCTGAAAAACGACCCA 861
501 laThrLeuValAsnIleSerProAspThrSerLeuLeuLeuAspCysGly 517
: : : ||| : : : : : |||
860 TAGCGCTTAAACACTACAGGTGCTGGG.....GGC 832
518 GluGlyThrPheGlnGlnLeuCysArgHisTyrglnGlyAspGlnValAspAr 534
||||| ||||| : : : : : |||||
831 ATTGACACACTCTGACAAAGTGTGCACTCACTCGCTCTCAG..... 790

534 gValLeuGlyThrLeuAlaAlaValPheValSerHisLeuHisAlaAspH 551
: : ||| ||||| : : : : : |||||
789GAGAGTCTGGACGCTGCACCTGTGAGGCGAGGACAC.....C 753
551 isHisThrGlyLeuProSer.....IleLeuLeuGlnArgGluArg... 564
||||| ||||| : : : : : |||||
752 ACCACGAGGAGCAGTGGTGTCTTCGTGTTCTGACGCGCATCATCGG 703
565AlaLeuAlaSerLeuGlyLysPro..... 572
||||| ||||| : : : : : |||||
702 AACCTGACTCTTGACGTGGCTGCATGTGGCTTCCGGTATTATTACAGCCT 653
573 .LeuHisProLeuLeuValValAlaProAsnGlnLeuLysAlaTrpLeuG 589
||||| : : : ||||| : : : : : |||||
652 TCTCCAGCTGTCTCGCAGGTCCTTTCTCAAGACCTTTT..... 614
589 lnGlnTyrglnHisAsnGlnCysGlnGluValLeuHisHisIleSerMetIle 605
: : : : : ||||| |||||
613CTTTGCACTCAGGACGACAC..... 593
606 ProAlaLysCysLeuGlnGluGlyAlaGlu...IleSerSerProAlaVa 621
: : : ||||| : : : ||||| : : : : : |||||
592 ...ATGGAAGTCTTCAAAATGGCAATCATTTTAAATGCACCAGCAGA 546
621 lGluArgLeuIleSerSerLeuLeuArgThrCys..... 632
: : : : : ||||| : : : |||||
545 TGTCGCCAGCATTAACCTGCTCGCACAAACCTCTGCTTTTCATTCGACAATA 496
633AspLeuGluGluPheGlnThrCys..... 640
: : : ||||| : : : ||||| : : : : : |||||
495 GATCTGAAGAGCCAGAAATTTCTCTTTCGACGAATATCTCTAAACACCT 446
641 ...LeuValArg..... 643
||| ||| |||||
445 TATCTTTAAGCATGCTCTCTTGACACGCTGTACATTTTGGACTTGAAGAC 396
644HisCysLysH 647
||| ||||| : : : : : |||||
395 CTCACGAGGCGCCCATGACGCTCTCGCAGAAGCGTGCCACACACCTCGGT 346
647 isAlaPheGlyCys.....AlaLeuValHisThrSerGlyTrp 659
||| ||||| : : : : : |||||
345 CTGCTCGGGCTGCACACGACGACGAGTGGGCTCTCTCACACTTGTACTGT 296
660 LysValValTyrglnSerGlyAspThrMetProCysGluAlaLeuValArgMe 676
||||| : : : : : ||||| : : : : : |||||
295 CCT...CCAGGCTCTCACAAACTTTCTTGTANACCTCTTGTTCAGGG 249
676 tGlyLysAspAlaThr 681
||| ||||| : : : : : |||||
248 ACAAAACTGGCGTCC 233

seq_name: /cqln_7/ptodata/1/lna/5B_COMB.seq:US-08-804-227C-7

seq_documentation_block:
; Sequence 7, Application US/08804227C
; Patent No. 5876991
; GENERAL INFORMATION:
; APPLICANT: Dehoff, Bradley S.
; APPLICANT: Kunstoss, Stuart A.
; APPLICANT: Rostock, Paul R., Jr.
; APPLICANT: Sutton, Kimberly L.
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THOMAS G. PLANT 1501
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:

```

: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: MS-DOS
: SOFTWARE: ASCII(DOS) Text only
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/804.227C
: FILING DATE: February 21, 1997
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Plant, Thomas, G.
: REGISTRATION NUMBER: 35,784
: REFERENCE/DOCKET NUMBER: X-8231
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 317-276-2459
: INFORMATION FOR SEQ ID NO: 7:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 44377 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 350..14002
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 14046..20036
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 20110..31284
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 31329..36071
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 36155..41830
: US-08-804-227C-7

```

```

alignment_scores:
  Quality: 113.50      Length: 922
  Ratio: 0.314        Gaps: 46
  Percent Similarity: 39.154  Percent Identity: 20.390

```

```

alignment_block:
US-09-434-382-2 x US-08-804-227C-7
Align seg 1/1 to: US-08-804-227C-7 from: 1 to: 44377

2 TrpAlaLeuCysSerLeuLeuArgSer...AlaAlaGlyArgThrMet... 16
|||
|||
|||
37032 TGCGGTGGTGGCGGCGAGTGGCGTCAATCAGGACGGCGGAGCAATGGG 37081
17 .....SerGlnGlyArgThrIleSerGlnA 25
|||
|||
|||
37082 CTGACGGCGCGGAGTGGTCTGCGCAGCAGCGGGTATCGTGAGCGGT 37131
25 laProAlaArgGluArgPro..... 32
|||
|||
|||
37132 GGCTGATCGGGGCTGACGCCGCCACGCTGGATGTGGTGAGGCGCAGC 37181
33 .....ArgLysAspProLeuArgHisLeuArgThrArgGluLysAr 46
|||
|||
|||
37182 GTACGGGAGCGGCGTGGGTGATCCGATCGAGCGGGTGGT..... 37223
46 gGlyProSerGlyCys.....SerGlyGlyProAsnThrValTyrL 60
|||
|||
|||
37224 .....TGTCGCCACGTACGGCGGAGCGGGTGGTATCC 37260
60 euGlnValValAlaAlaGlySerArgAspSerGly.....AlaAlaLeu 74
|||
|||
|||
37261 TTGTGGTGGGTGCTTGANGTCGACATCGGGCATGCCAGCGGGCTG 37310

```

```

75 TyrValPheSerGluPheAsnArgTyrLeuPheAsnCysGlyGluGlyVa 91
|||
|||
|||
37311 CGGTGTGGTGGTGTGATCAAGGTGTGAGGCGATCCGCGATGGTGC 37360
91 lGlnArg.....LeuMetGlnGluHisLysLeuLysValAlaArgL 105
|||
|||
|||
37361 TTGCCGCGGAGCGTGTGATGGATGCCCGTCTGTCGAGAGGTGGAGTGGC 37410
105 euAspAsnIlePheLeuThrArgMetHisTrpSer..... 116
|||
|||
|||
37411 TT.....CGGTGGGTGGAGCTGCTGACCGAGGGGCC 37442
117 .....AsnValGlyGlyLeuSerGlyMetIleLeuThrLeuLysGluTh 131
|||
|||
|||
37443 GGTCGTGGCGCGCGGTGGAGCGG.....TG 37471
131 rGlyLeuProLysCysValLeuSer.....GlyProGlnLeuG 145
|||
|||
|||
37472 CGCGGGCGCGGTGTGCGGTTCGGGTGAGCGGACCAACGCCCATGT 37521
145 lulystyrLeuGluAlaIleLysIlePheSerGlyProLeuLysGlyLe 161
|||
|||
|||
37522 GGTCTCTGG..... 37529
162 GluLeuAlaValArgProHis.....Se 169
|||
|||
|||
37530 .....AGGAAGCACCGGTGAGCGCGGAGCAGCAGCGGAC 37567
169 rAlaProGluTyrGluAspGluThrMetThrValTyrGlnIleProIleH 186
|||
|||
|||
37568 GGCCCCGGACCGCACCGCGCGAGCGGTGACGGGTCCGTCCCTGGGT 37617
186 isSerGluGlnArgArgGlyLysHis.....GlnProTyrGlnSer 199
|||
|||
|||
37618 GCTCTCGGCACGCTCGCGGGAGCGGTGCGCGGCCAGCGCGGAGCTCG 37667
200 ProGluArgProLeu.....SerArgLeuSerProGluArgSerSerAs 214
|||
|||
|||
37668 CCGCTCTCGCGCGCGCAGGGCGCAGGAGGCGCACCGCGCGGCGGCGGA 37717
214 pSerGluSer..... 217
|||
|||
|||
37718 CTGCTCTCGCGCGCGCGCATCGGATCTCCTTGGCCACCAACAGGGA 37767
217 ..... 217
37768 GACCCTGGAGCACCGCGCGGTGGCTGTCAGGAGAACCGGACGGCGCG 37817
218 AsnGluAsnGluProHisLeuProHisGlyValSerGlnArgGlyVa 234
|||
|||
|||
37818 GGGAGGACCTCGCGCGCTGCGCGCGCGCGCACACCGGAGAGCGTGTGC 37867
234 lArgAspSerSerLeuValValAlaPheIleCysLysLeuHisLeuLysA 251
|||
|||
|||
37868 ACGGGTCTCGCGACGTGCGCGG.....GGATCGCTTCTCTCTGCTC 37911
251 rgGlyAsnPheLeuValLeuLysAlaLysGluMetGlyLeuProValGly 267
|||
|||
|||
37912 GGGGC.....AGGCGCGCCAGCGGTGCGCGCGCGGTGGG 37946
268 ThrAlaIleLeuAla...ProIleIleAlaValLysAsp..... 280
|||
|||
|||
37947 AGCTCCCGCGCAGGTCTCCCGTCTTCGCGCAGCGCTCGACGAGATCGCG 37996
281 GlyLysSerIleThrHisGluGlyArgGluIleLeuAlaGluGluLeuCy 297
|||
|||
|||
37997 GCGGAGTTCAGCGCCACCTCGAACGCCCTCTCTCTCGGTGATGTTCG 38046
297 sThrProAspProGlyAlaAlaPheValValGluCysProAspG 314
|||
|||
|||
38047 CGACCGCGCGCGGAGCGCGCCTCTCTCTCTCTCTCTCTCTCTCTCTCT 38084
314 luSerPheIleGlnPro.....IleCysGluAsnAlaThrPheGln 327

```

```
38085  CGACTACACCGAGCGCGCTTTCGCGGTGGAGACCGCGCTCTTCGG 38134
328  ArgTyrGln.....
38135  CTCCTGGAGAGCTGGGGCTGGTCCCGGACGTCTCGTGGGGACACTCGAT 38184
331  ....GlyLysAlaAspAlaProValAlaLeuValValHisMetAlaProA 346
38185  CGGCGGTCTGGTGGCGGTACAGTGGCGGCGTCTCTCTGCGGCGGACG 38234
346  laSerValLeuValAspSerArgTyrGln.....
38235  CGGCCGCTGTCTCTCCGACGCGCGGCTCATCGGGGCCCTGCCGAG 38284
356  ....GlnTrpMetGluArg.....
38285  GCGCGCGGATGGCGCGGTGACGCCACCGAGCGGAGCGCGCGCT 38334
361  ....PheGlyProA 364
38335  GGAGCCGTCGCGCGCGGCGGCGGTGTGTCGCGCGGTCAACGCGCG. 38383
364  spThrGlnHisLeuValLeuAsnGluAsnCysAlaSerVal.....
38384  ....CAGGCCCTGTCTCTCCGGGAGCGAGCGCGGTACTGCGGCG 38428
378  ....HisAsnLeuArgSerHisLysIleGlnThrG1 388
38429  GCCGTGTAAGTGGCGCGCGGACCGCGCACCAAGCGCTGAGGGTGAG 38478
388  nLeuAsnLeuIleHisProAspIlePheProLeuLeuThrSerPheArgC 405
38479  CCAGCCCTTCCACTCACCCGATGAGCGCATGCTCGCGGATTCGCG 38528
405  ysLysLysGlu.....GlyProThrLeuSer...ValPro 415
38529  CGGTGGCGGACACGCTCGACTACACGCGCGCGCTCGCGTCTCTCC 38578
416  MetValGlnGlyGluCysLeuLeuLysTyrGlnLeuArgProArgG1 432
38579  GAAGTACCGGCGGACTCGCGACCGCGCGCGCTGACCGACCGCGGTA 38628
432  uTrpGlnArgAspAlaIleIleThrCysAsnProGluGluPheIleValG 449
38629  CTGGACCCGC.....C 38639
449  luAlaLeuGlnLeuProAsnPheGlnGlnSerValGlnGluTyrArgArg 465
38640  AGGTGCGGACCGGCTGCGCTTCGCGGACGCGCGTGC. ....ACCGCG 38683
466  SerAlaGlnAspGlyProAlaProAlaGluLysArgSerGlnTyrProG1 482
38684  AGCGCCCGGAC..... 38695
482  uIleIlePheLeuGlyThrGlySerAlaIleProMetLysIleArgAsnV 499
38695  .... 38695
499  alSerAlaThrLeuValAsnIleSerProAspThrSerLeuLeuLeuAsp 515
38696  ..GCGCGACCTTCATCGAGCTCGGCGCGGCTG..... 38734
516  CysGly.....GluGlyThrPheG1 522
38735  TCGGCGATGGCGGAGGATCCTTGGCCCGGAGCGCGGCTGTTCG 38784
522  yGlnLeuCysArgHisTyrGlyAspGlnValAspArgValLeuGlyThrL 539
38785  CCGCGCACTGCGGCGCGCGCGGAGCGGACACCGTGTCTCGGCGCG 38834
539  euAlaAlaValPheValSerHisLeuHisAlaAsp..... 550
|||||
```

```
38835  CGCGGAGCGGTACGTCCGCGCGCGGCTCGACTGGCGCGCGCTCTAC 38884
551  .....HisHisThrGlyLeuProSerIleLeuLeuGlnAr 562
38885  GCGCGCACGGGAGCGCGCGCACCTGCGCCACCTAGCCCTTCAGCA 38934
562  gGluArgAla...LeuAlaSerLeuGlyLysProLeuHisProLeuLeuV 578
38935  CAGCGCTACTGGTTCGCGCGCGCTCGCGCGCGGTGCGCCCC...GCCA 38981
578  alValAlaProAsnGlnLeuLysAlaIleTrpLeuGlnGlnTyrHisAsnGln 594
38982  CGCGCGCGCGCGCTCC..... 38995
595  CysGlnGlnValLeuHisHisIleSerMetIleProAlaLysCysLeuG1 611
38996  .....GTCCGATCCGTGCGG...GAAGCCGAGCA 39021
611  nGluGlyAlaGluIleSerSerProAlaValGluArgLeuIleSerSerL 628
39022  GGACGGGCGGTG..... 39034
628  euLeuArgThrCysAspLeuGluGluPheGlnThrCysLeuValArgHis 644
39034  ..... 39034
645  CysLysHisAlaPheGlyCysAlaLeuValHisThrSerGlyTriLysVa 661
39035  .....TGGCGCGC 39042
661  lValTyrSerGlyAspThrMetProCysGluAlaLeuValArgMetGlyL 678
39043  CGTGCACGCGGTGACGTC.....GCCTCGCGCGCGCGCGCTGGCG 39086
678  ysAsp.....AlaThrLeuLeuIleHisGlu 686
39087  CGGAGCGCGCGGTATCGAACACGAACCTGCGCGCGGTCTCGCGCACCTG 39136
687  AlaThrLeuGluAspGlyLeuGluGluAlaValGluLysThr..... 701
39137  GCGCGCTGGCACGAC.....CGGACCGCGCGCGCGCGCGCGCGG 39180
702  .....HisSerThrThrSerGlnAlaIle.....SerValGlyM 713
39181  CTGCACCTACCGCGTCACCTGCGCGCGGTATCGAGCAGACGCTGTCAGGT 39230
713  eArgMetAsnAlaGluPheIleMetLeuAsnHis.PheSerGlnArgTy 729
39231  TCAGCCCTCGGATCGCTGGTGTGATGGTATGGCATGGGCGACACGGAA 39280
729  rAlaLysValPro 733
39281  TCGCGGAGCGCG 39293
seq_name: /cgnl_7/ptodata/1/ina/5B_COMB.seq:US-08-804-198-1
seq_documentation_block:
: Sequence 1, Application US/08804198
: Patent No. 5945320
: GENERAL INFORMATION:
: APPLICANT: Burgett, Stanley G.
: APPLICANT: Kuhstoss, Stuart A.
: APPLICANT: Rao, Nagaraja R.
: APPLICANT: Richardson, Mark A.
: APPLICANT: Rosteck, Paul R., Jr.
: TITLE OF INVENTION: PLATENOLIDE SYNTHASE GENE
: NUMBER OF SEQUENCES: 6
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: PAUL R. CANTRELL 1138
: STREET: LILLY CORPORATE CENTER
: CITY: INDIANAPOLIS
: STATE: IN
: COUNTRY: USA
```



```
314 luSerPheIleGlnPro.....IleCysGluAsnAlaThrPheGln 327
      ::::: |||||
38085 CGACTACACCGCGCGCCCTTTCGGCGTGGAGACCGCGCTCTCCGG 38134
      ::::: |||||
328 ArgTyrGln..... 330
      ::::: |||||
38135 CTCCTGGAGAGTGGGGCTGTGTCGGGACGTCCTCGTGGGCGACATCGAT 38184
      ::::: |||||
331 ....GlyLysAlaAspAlaProValAlaLeuValHisMetAlaProA 346
      ::::: |||||
38185 CGCGGTCTGGTGGCGGCTCAGCTCGGGGCGTCTCTCTGGCGCGGACG 38234
      ::::: |||||
346 laSerValLeuValAspSerArgTyrGln..... 355
      ::::: |||||
38235 CGCGCGGCTGTCTCCGACGCGCGGCTCATGCGGGCCCTGTCGCGGAG 38284
      ::::: |||||
356 .....GlnTrpMetGluArg..... 360
      ::::: |||||
38285 GCGCGGCGATGCGGCGCTGTCAGGCCACCGAGCGGAGCGCCCGCT 38334
      ::::: |||||
361 .....PheGlyProA 364
      ::::: |||||
38335 GGAGCCGTCGCGCGCGCGGCGGTGTGTCGCGCGGTCAACGCCCG. 38383
      ::::: |||||
364 spThrGlnHisLeuValLeuAsnGluAsnCysAlaSerVal..... 377
      ::::: |||||
38384 .....CAGGCCCTCGTCTCTCGGGAGGAGCGGCGCTACTGCGCGG 38428
      ::::: |||||
378 .....HisAsnLeuArgSerHisLysIleGlnThrCl 388
      ::::: |||||
38429 GCGGTGAAGTGGCGCGCGCGCGCGCGCCACCAACCGCTGAGGGTGAG 38478
      ::::: |||||
388 nLeuAsnLeuIleHisProAspIlePheProLeuLeuThrSerPheArgC 405
      ::::: |||||
38479 CCACGCCCTCCACTCACCCGTATGAGCCCATGCTCGCCGACTTCCGG 38528
      ::::: |||||
405 ysLysLysGlu.....GlyProThrLeuSer...ValPro 415
      ::::: |||||
38529 CGGTGGCGGACACGCTCGACTACCAACGCGCGCGCGCTGCGGTCTCTCC 38578
      ::::: |||||
416 MetValGlnGluCysLeuLeuLysTyrGlnLeuArgProArgArgGl 432
      ::::: |||||
38579 GAAGTACCGCGGCGCTCGCCAGCGCGCGCGCTGACCGACCGCGGTA 38628
      ::::: |||||
432 uTrpGlnArgAspAlaIleIleThrCysAsnProGluGluPheIleValG 449
      ::::: |||||
38629 CTGGACCGC.....C 38639
      ::::: |||||
449 luAlaLeuGlnLeuProAsnPheGlnSerValGlnGluTyrArgArg 465
      ::::: |||||
38640 AGGTGCGGACGCGGTGGCTTCGCGCGACGCGCGTGC.....ACCGCG 38683
      ::::: |||||
466 SerAlaGlnAspGlyProAlaProAlaGluLysArgSerGlnTyrProGl 482
      ::::: |||||
38684 ACCGCGCGGAC..... 38695
      ::::: |||||
482 uIleIlePheLeuGlyThrGlySerAlaIleProMetLysIleArgAsnV 499
      ::::: |||||
38695 ..... 38695
      ::::: |||||
499 alSerAlaThrLeuValAsnIleSerProAspThrSerLeuLeuLeuAsp 515
      ::::: |||||
38696 ..GCCGCGACCTTCATGAGCTCGGCGCGCGCGCTGCTC..... 38734
      ::::: |||||
516 CysGly.....GluGlyThrPheGl 522
      ::::: |||||
38735 TCGCGCATGGCGGAGAGTCCCTGCGCGGAGCGCGCGCTGCTCGC 38784
      ::::: |||||
522 yGlnLeuCysArgHisTyrGlyAspGlnValAspArgValLeuGlyThrL 539
      ::::: |||||
38785 CCGCGCACTGGCGCGCGCGCGCGGAGCGGACACCGTGTCTCCGGCGG 38834
      ::::: |||||
```

```
539 euAlaAlaValPheValSerHisLeuHisAlaAsp..... 550
      ::::: |||||
38835 CGCGAGCGGTACGTCCGCGCGCGGCTCGACTGGCGCGCGCTCTAC 38884
      ::::: |||||
551 .....HisHisThrGlyLeuProSerIleLeuLeuGlnAr 562
      ::::: |||||
38885 GCGCGCACGGGACCGCGCGCGCGCTGCGCGCGCTACGCTTCCAGCA 38934
      ::::: |||||
562 gGluArgAla...LeuAlaSerLeuGlyLysProLeuHisProLeuLeuV 578
      ::::: |||||
38935 CAGCGCTACTGGCTCGCGCGCGCGCTGCGCGCGCTGCGCGCG...CGCA 38981
      ::::: |||||
578 alValAlaProAsnGlnLeuLysAlaTrpLeuGlnGlnTyrHisAsnGln 594
      ::::: |||||
38982 CGCGCGCGCGCTCC..... 38995
      ::::: |||||
595 CysGlnGluValLeuHisHisIleSerMetIleProAlaLysCysLeuGl 611
      ::::: |||||
38996 .....GTCCGATCCGTGCCG...GAAGCCGAGCA 39021
      ::::: |||||
611 nGluGlyAlaGluIleSerSerProAlaValGluArgLeuIleSerSerL 628
      ::::: |||||
39022 GGACGGGCGCTG..... 39034
      ::::: |||||
628 euLeuArgThrCysAspLeuGluGluPheGlnThrCysLeuValArgHis 644
      ::::: |||||
39034 ..... 39034
      ::::: |||||
645 CysLysHisAlaPheGlyCysAlaLeuValHisThrSerGlyTrpLysVa 661
      ::::: |||||
39035 .....TGGGCCG 39042
      ::::: |||||
661 lValTyrSerGlyAspThrMetProCysGluAlaLeuValArgMetGlyL 678
      ::::: |||||
39043 CGTGACGCGCGGTGAGTC.....GCCTGCGCGCGCGCGCGCTGGCGG 39086
      ::::: |||||
678 ysAsp.....AlaThrLeuLeuIleHisGlu 686
      ::::: |||||
39087 CGGACGACGCGCGGTATCGAACACGAACTGCGCGCGGTCTCGCGCACCTG 39136
      ::::: |||||
687 AlaThrLeuAspGlyLeuGluGluAlaValGluLysThr..... 701
      ::::: |||||
39137 GCGCGCTGGCAGC.....CGCAGCGCGCGCGCGCGCGCGCGCGG 39180
      ::::: |||||
702 .....HisSerThrThrSerGlnAlaIle.....SerValGlyM 713
      ::::: |||||
39181 CTGCACTACCGCTCACCTGGCAGCGATCGAGCGAGCGCTGTCACT 39230
      ::::: |||||
713 etArgMetAsnAlaGluPheIleMetLeuAsnHis.PheSerGlnArgTy 729
      ::::: |||||
39231 TCAGCCCTCGGATCGCTGGTGTGATGTCGAGCATGGGCGAGCACCGAA 39280
      ::::: |||||
729 rAlaLysValPro 733
      ::::: |||||
39281 TCGCGGAGCGCG 39293
      ::::: |||||
seq_name: /cgnl_7/ptodata/1/ina/6_COMB.seq:US-09-320-878-21
seq_documentation_block:
; Sequence 21, Application US/09320878A
; Patent No. 6117659
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 300622002120
; CURRENT APPLICATION NUMBER: US/09/320,878A
; CURRENT FILING DATE: 1999-05-27
; EARLIER APPLICATION NUMBER: CIP OF 09/141,908
; EARLIER FILING DATE: 1998-08-28
```



```

; EARLIER APPLICATION NUMBER: CIP OF 09/073,538
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: CIP OF 08/846,247
; EARLIER FILING DATE: 1997-04-30
; EARLIER APPLICATION NUMBER: 60/119,139
; EARLIER FILING DATE: 1999-02-08
; EARLIER APPLICATION NUMBER: 60/100,880
; EARLIER FILING DATE: 1998-09-22
; EARLIER APPLICATION NUMBER: 60/087,080
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 21
; LENGTH: 5970
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-320-878-21

alignment_scores:
    Quality: 113.00      Length: 899
    Ratio: 0.298         Gaps: 43
    Percent Similarity: 42.158      Percent Identity: 19.800

alignment_block:
US-09-434-382-2 x US-09-320-878-21
..
Align seg 1/1 to: US-09-320-878-21 from: 1 to: 5970

      4  LeuCysSerLeuLeuArgSerAlaAla.....GlyArgThrMetSerG1 18
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2955  CTCTGCCAGTTACTCCGACGCGCGGGAACACACACGCGTCCGACACCTCCGC 3004
      |||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
18  nGlyArgThrIleSerGlnAlaProAlaArgArg.....GluArgProA 33
      |||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
3005  CGTCGCGGACCTGCTGACGCGGCTCGCGGCGGCGGCGGAACCTCACCTC 3054
      |||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
33  rGlyAspProLeuArgHisLeuArgThrArgGluLysArgGlyProSer 49
      |||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
3055  CCCGCGGACGCGCGCTCGCGCTCGACGACACACCGCGCTGCACGACT 3104
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
50  GlyCysSerGlyGlyPro...AsnThrValTyrLeuGlnValAlaAla 65
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
3105  GGCTGCTCTCCGCTCCGCGCGGCGGCGGCGGCTCTCTCTCGCGCGGACGA 3154
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
65  aGlySerArgAspSerGlyAlaAlaLeuTyrValPheSerGluPheAsnA 82
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
3155  CCTGACCCACGCGACACCGCGTCCCTG.....A 3183
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
82  rGlyLeuPheAsnCys.....GlyGluGly... 90
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
3184  GGTCTCTCTGTACTGCGCGCGGCCACACACGCGGCGGCGGCGGCTTC 3233
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
91  .....ValGlnArgLeuMetGlnGluHisLysLeuLysValAlaArgLe 105
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
3234  GTCATGACCGAGCGGCTCGCGCGCGCGCGGATACCGGGGTTCGCGCG 3283
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
105  uAspAsnIlePheLeuThrArgMetHisTrpSerAsn.....ValGlyG 120
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
3284  CGAG.....CTGCTCGCGCGCGGCGGCTGCGCGCAACATGTGGCTCTCCG 3327
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
120  lYLeu.....SerGlyMetIleLeuThrLeuLysGlu.ThrGlyLeuPr 134
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
3328  GGCTTCCCGCGCGGCGGTACGCGAGTTACTCGGCCACTACTAGCGGCC 3377
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
134  oLysCysValLeuSerGlyProGlnLeuGluLysTyrLeuAlaAla 151
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
3378  GAGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGAA 3427
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
151  leLysIlePheSerGly.....ProLeuLysGlyIleGluLeuAlaVal 165
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
3428  CCCGCTGCTCTCGCGGCGGCTGACCCAGGACCGGC..... 3462

```

```

166  ArgProHisSerAlaProGlu.....TyrGluAspGluTh 177
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
3463  AGGCTCCACACACACCTCGCGCGCGCGCGCGCGAGCGCGTCCAC 3512
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
177  rMetThrValTyrGlnIleProIleHisSerGluGlnArgArgGlyLysH 194
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
3513  GGCACGCGCTTCGCGCGCGCGCTCTGCTGCTGCTGCTGCTGCTGCTGCTG 3562
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
194  iGlnProTrpTrpGlnSerProGluArgProLeuSerArgLeuSerProGlu 210
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
3563  GGCACACACTGGAGACCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3612
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
211  ArgSer.....SerAspSerGluSerAsnGluAsnGluPro.HisLeuP 225
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
3613  CGCTCTCTGGTGGAGCGCTCACGGGAACGACCGCGCGCGCTGAGCGC 3662
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
225  roHisGlyValSerGlnArgArg.GlyValArgAspSerSerLeuValVa 241
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
3663  CACATCCAGGAGCTCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3712
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
241  lAlaPheIleCysLysLeuHisLeuLysArgGlyAsnPheLeuValLeuL 258
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
3713  G.....GACAGCGCGCGATCCGCG 3732
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
258  ysAlaLysGluMetGlyLeuProValGlyThrAlaAlaIleAlaProIle 274
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
3733  AGGCGCGCTCCAGGAGCTCGCGCGCGGAGCGACCGCACTGCACCGG 3782
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
275  IleAlaAlaVal.....LysAspGlyLysSerIleThrHisGluG1 288
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
3783  CGCGCGCGGAGCGTGCACCGGCGCGCGCGCGCGAGGACCGCGTGGC 3832
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
288  yArgGluIleLeuAlaGluGluLeuLysCysThrProAspProGlyAlaA 305
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
3833  CGCGCACTGCTGCTC.....GCGGCG 3855
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
305  laPheValValGluCysProAspGluSerPheIleGlnProIleCys 321
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
3856  CC.....CCCGACGCTCCTGCGCGCTGCGCGCTGCTGCTGCTGCTGCT 3887
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
322  GluAsnAlaThrPheGlnArgTyrGlnGlyLysAlaAspAlaProValAl 338
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
3888  GAACGGGCGCGGACGAC..... 3905
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
338  aLeuValValHisMetalProAlaSerValLeuValAspSerArgTyrG 355
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
3906  .....GCCCTGTTCGACGCGCGACGCTC. 3926
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
355  lnGlnTrpMetGluArgPheGlyProAspThrGlnHisLeuValLeuAsn 371
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
3927  .....GACGACGCTTCCGGATCTCGAG 3950
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
372  GluAsnCysAlaSerValHisAsnLeuArgSerHisLysIleGlnThrG1 388
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
3951  .....TTCGCGCTGCGGTCGAGACCGCGACACACCA 3982
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
388  nLeuAsnLeuIleHisProAspIlePheProLeuLeuThrSerPheArg 405
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
3983  GCTGCGCGCGCTCCCGCCACACTG.....GTGCGCGCTCTCTGGCGGA 4026
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
405  ysLysLysGluGlyProThrLeuSerValPrometValGlnGlyCys 421
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
4027  TGAACCGGCACATGACGACCGCGCGCTGCTGCTGCTGCTGCTGCTGCTG 4070
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
422  LeuLeuLysTyrGlnLeuArgProArgGluTrpGlnArgAspAlaI1 438
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
4071  CTCCTGAGCGGTGAAGTCCCGCGCGCGCGCGCGCGCGCTCATG.....GCCCT 4114
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
438  eIleThrCysAsnProGluGluPheIleValGluAlaLeuGlnLeuProA 455
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
4115  GATCCGCTGC.....CTCGCTGGTACGCGCGCTGCGCG 4149
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
455  snPheGlnGlnSerValGlnGluTyrArgArgSerAlaGlnAspGlyPro 471

```

```
.. .. . ||| |||:.....
4150 AGCCGCCGACCGCTGTCCCGCTGGCGCCAGCTCCCGACACGATGCC 4199
472 AlaProAlaGluLysArgSerGlnTyrProGluIleIlePheLeuGlyTh 488
4200 TTGGAGCTG.....TCGCTCACCCGGATGGCT 4228
488 rGlySerAlaIleProMetLysIleArgAsnValSerAlaThr..... 502
4229 CGCGCGCTGTCCCGCCGCTCTCTGGAGTCCCTGCGCGCACGCCGAGC 4278
503 .....LeuValAsnIleSerProAspThrSerLeuLeu 513
4279 CGAGGGGGTCCGTCCTCCCGTACCGCTCGCGCCGGGACGCC..... 4322
514 LeuAspCysGlyGluGlyThrPheGlyGlnLeuCysArgHisTyrGlyAs 530
4322 ..... 4322
530 pGlnValAspArgValLeuGlyThrLeuAlaAlaValPheValSerHisL 547
4323 .....GGCTCCAGCCCGACGCGCGCTCTTC..... 4349
547 euHisAlaAspHisThrGlyLeuProSerIleLeuLeuGlnArgGlu 563
4350 .....CAGCGGGGC 4358
564 ArgAlaLeuAlaSerLeuGlyLysProLeuHisProLeuValValAl 580
4359 CCGGACACGCCCTCGTCCG..... 4379
580 aProAsnGlnLeuLysAlaTrpLeuGlnTyrHisAsnGlnCysGlnG 597
4380 .....CAGGCGGAAC 4389
597 luValLeuHisHisIleSerMetIleProAlaLysCysLeuGlnGly 613
4390 AGATCTG.....CAGGGC 4403
614 AlaGluIleSerSerProAlaValGluArgLeuIleSerSerLeuAr 630
4404 TCCCGGCTGTCCGAGGAGAGCTACGAGGCCCTGGAGACGGCCCTCTGT 4453
630 gThrCysAspLeuGluGluPheGlnThrCysLeuValArgHisCysLysH 647
4454 CCGCTCCACGCGACCGGCTCGACCGGCGCTGTC..... 4490
647 isAlaPheGlyCysAlaLeuVal.....HisThrSer 657
4491 .....TGCTCGGACGCCCTCTCGCGGAGCGCTGGAGCGGCGTCTC 4535
658 GlyTrpLysValValTyrSerGlyAspThrMetProCysGluAlaLeuVa 674
4536 GCCTGGAGCGGCTCTCGCGCGACCCGG..... 4565
674 lArgMetGlyLysAspAlaThrLeuLeuIleHisGluAlaThrLeuGluA 691
4566 .....GGCATCGCATCGCTCGCGGACCTCCCGA 4599
691 spGlyLeuGluGluAlaValGluLysThrHisSerThr...ThrSer 706
4600 CGGCGGGAGCGGCGGAGCTGGCGCTCTCCACGCGCGCGGAGAGC 4649
707 GlnAlaIleSerValGlyMetArgMetAsnAlaGluPheIleMetLeuAs 723
4650 TGGGGCCTCGCGTGGCATGCCCTCTCCGGCTGCTGCTCGCTCGCTGCAC 4699
723 nHisPhe.....SerGlnArgTyrAlaLysValProLeuP 735
4700 GGAGCGCGCGGAGTACGAACAGCGGAGCGGCTCTCGCGGACGCGGTGC 4749
735 heSerProAsnPheSerGluLysValGlyValAlaPheAspHis..... 749
```

```
4750 CGGACCGCATGTTCGACTCGCGGCACGGCATGGAGTACATGCACGCCCGG 4799
750 .....MetLysValCysPheGlyAspPh 757
4800 GGCCGCTACTGGCTGGCGANCGCGGCTGCACGCGCGCTGGCGAGTT 4849
757 eProThrMetProLysLeu.....I 764
4850 CATGCTCTCGGGGAGATCTGGGCAGCTGGAACCTCGACACGCCCTCGA 4899
764 leProProLeuLysAlaLeuPheAlaGlyAspIleGluMetGluGlu 780
4900 TCGTCCCTGGCGGAGCTCCGCGCGGAGGTGTACCTCGGCTCGGCAAC 4949
781 ArgArgGluLysArgGluLeuArgGlnValArgAlaAlaLeuLeuSer... 796
4950 CGCCAGAGCCAGGCGCTGGCCGAGGCCACCTCGCCCTGGTGGCGGC 4999
797 .....ArgGluLeuAlaGlyGlyL 803
5000 CGGGGCGCTCCCGCACCCGGGCTCTACCCCTGGGGTCTTGGCGCGCGG 5049
803 euGluAspGlyGluProGlnGlnLysArgAlaHisThrGlu 816
5050 TG...GACGGCCAG...CAGCGGAGCGGCTGCACGCCGAG 5084
```

OM of: US-09-434-382-2 to: EST:* out_format : pfs

Date: Feb 18, 2001 4:09 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODE=frame+p2n.model -DEV=xlp
-O=/cgn1_1/USPTO_spool/US09434382/runat_16022001_105748_29671/app_query.fasta_1.895
-DB=EST -QFMT=fastap -SUFFIX=rst -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000 -GAPOP=4.500
-GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
-ALIGN=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
-MAXLEN=200000000 -USER=US09434382.cgn1_1_4002 -NCPU=6
-ICPU=3 -LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-434-382-2

Query length: 826

Database: EST:*

Database sequences: 7991742

Database length: -791223438

Search time (sec): 1395.210000

score_list:

Sequence	Strd Orig	Zscore	EScore	Len	Documentation
gb_est75:BE795820	+ 1242.50	2207.27	6.2e-114	761	BE795820 601590856F1 NIH_MGC_7
gb_est69:BE260495	+ 1180.00	2096.49	9.2e-108	676	BE260495 601150702F1 NIH_MGC_19
gb_est75:BE744197	+ 1165.00	2065.13	5.1e-106	949	BE744197 601577168F1 NIH_MGC_9
gb_est73:BE619259	+ 1164.50	2066.08	4.5e-106	823	BE619259 601473130F1 NIH_MGC_68
gb_est75:BE744876	+ 1158.50	2052.98	2.4e-105	984	BE744876 601576324F1 NIH_MGC_9
gb_est76:BE867512	+ 1134.50	2048.28	4.4e-105	812	BE867512 601443010F1 NIH_MGC_9
gb_est77:BE794311	+ 1131.00	2019.63	1.8e-103	992	BE794311 601577254F1 NIH_MGC_9
gb_est77:BE892893	+ 1123.00	1992.03	6.0e-102	790	BE892893 601435738F1 NIH_MGC_7
gb_est77:BE900936	+ 1109.00	1967.39	1.1e-101	745	BE900936 601674206F1 NIH_MGC_21
gb_est77:BE902696	+ 1086.50	1928.68	2.0e-98	664	BE902696 601677349F1 NIH_MGC_21
gb_est70:BE383336	+ 1066.00	1891.30	2.5e-96	692	BE383336 601298249F1 NIH_MGC_19
gb_est69:BE260626	+ 1005.00	1777.73	5.2e-90	938	BE260626 601146116F1 NIH_MGC_19
gb_est75:BE742908	+ 1000.50	1773.87	8.6e-90	677	BE742908 601574609F1 NIH_MGC_9
gb_est69:BE304720	+ 996.00	1767.09	2.0e-89	612	BE304720 601106236F1 NIH_MGC_15
gb_est75:BE743831	+ 977.00	1727.38	3.3e-87	940	BE743831 601577742F1 NIH_MGC_9
gb_est72:BE537086	+ 975.00	1727.37	3.3e-87	713	BE537086 601063325F1 NIH_MGC_10
gb_est3:AA184645	+ 935.00	1657.98	2.4e-83	588	AA184645 mt58d10.r1 Soares thym
gb_est4:AA243700	+ 897.00	1591.01	1.3e-79	531	AA243700 zr68q08.sl Soares NIH
gb_est68:BE250309	+ 897.00	1590.00	1.5e-79	574	BE250309 600943455F1 NIH_MGC_17
gb_est41:AW206103	+ 892.00	1581.91	4.2e-79	536	AW206103 UI-H-B11-afx-f12-0-UT
gb_est77:BE914535	+ 882.00	1562.11	5.3e-78	617	BE914535 601665633F1 NCI_CGAP_M
gb_est49:AW771657	+ 871.00	1544.79	4.9e-77	511	AW771657 hn59h06.x1 NCI_CGAP_M
gb_est46:AW575677	+ 857.00	1520.09	1.2e-75	493	AW575677 UI-HF-BM0-adl-b-07-0-UT
gb_est70:BE409312	+ 855.50	1513.02	2.9e-75	691	BE409312 601300940F1 NIH_MGC_21
gb_est37:AV615441	+ 841.00	1490.08	5.5e-74	543	AV615441 AV615441 Bos taurus ad
gb_est70:BE371842	+ 828.00	1461.65	2.1e-72	803	BE371842 601217643F1 NCI_CGAP_M
gb_est73:BE615669	+ 825.00	1455.19	4.8e-72	872	BE615669 601279347F1 NIH_MGC_39
gb_est41:AW175923	+ 818.00	1449.51	1.0e-71	512	AW175923 RCL-BT0077-020999-002
gb_est7:AA455121	+ 809.50	1434.72	6.7e-71	493	AA455121 zx78c04.sl Soares ovar
gb_est5:AA311855	+ 809.00	1433.62	7.7e-71	501	AA311855 EST182568 Jurkat T-cell
gb_est4:AA291670	+ 807.00	1430.74	1.1e-70	474	AA291670 zt37d04.sl Soares ovar
gb_est17:AI200296	+ 788.00	1397.24	8.1e-69	451	AI200296 qf86b12.x1 Soares feta
gb_est10:AA531869	+ 771.50	1364.37	5.5e-67	578	AA531869 v105g04.r1 Barstead md
gb_est8:AA634909	+ 762.00	1350.66	3.2e-66	446	AA634909 ab27h02.r1 Barstead md
gb_est22:AI553139	+ 760.00	1345.37	6.3e-66	508	AI553139 v105g04.r1 Barstead md
gb_est5:AA310236	+ 754.00	1336.72	1.9e-65	431	AA310236 EST181085 Jurkat T-cell
gb_est10:AA679618	+ 741.00	1312.74	4.1e-64	452	AA679618 ag72c12.sl Gessler Wil
gb_est7:AA409793	+ 740.00	1311.44	4.9e-64	435	AA409793 EST015133 Mouse 7.5 opd
gb_est72:BE540045	+ 735.50	1295.20	3.9e-63	816	BE540045 601060822F2 NIH_MGC_10

gb_est86:W27286 - 735.00 1293.70 4.8e-63 855 ! W27286 28h1 Human retina cDN
gb_est20:AI468143 + 734.00 1301.08 1.8e-63 421 ! AI468143 tf92g05.x1 NCI_CGAP
gb_est84:R87541 + 730.50 1292.11 5.8e-63 518 ! R87541 ym9b04.r1 Soares adu
gb_est42:AW304130 + 702.00 1244.11 2.8e-60 404 ! AW304130 xs13e05.x1 NCI_CGAP
seq_name: gb_est75:BE795820
seq_documentation_block:
LOCUS BE795820 761 bp mRNA EST 20-SEP-2000
DEFINITION 601590856F1 NIH_MGC_7 Homo sapiens cDNA clone IMAGE:3945085 5',
mRNA sequence.
ACCESSION BE795820
VERSION BE795820.1 GI:10217018
SOURCE EST.
KEYWORDS human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 761)
NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: DCTD/DTF
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LLC802 row: p column: 14
High quality sequence stop: 761.
Location/Qualifiers
1..761
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3945085"
/clone_lib="NIH_MGC_7"
/tissue_type="small cell carcinoma"
/cell_line="MGC3"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lung; Vector: pOTB7; Site:1: XhoI; Site:2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5',
adaptor: GGACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 172 a 206 c 219 g 164 t
ORIGIN

alignment_scores:

Quality: 1242.50 Length: 253
Ratio: 5.010 Gaps: 1
Percent Similarity: 98.024 Percent Identity: 96.838

alignment_block:

US-09-434-382-2 x BE795820 ..

Align seg 1/1 to: BE795820 from: 1 to: 761

494 MetLysIleArgAsnValSerAlaThrLeuValAsnIleSerProAspThr 510

|||||

2 ATGAAGATTCGAATGTGTCAGGCCACACTTGTCAACATTAAGCCGACAC 51

|||||

510 rSerLeuLeuLeuAspCysGlyGluGlyThrPheGlyGlnLeuCysArgH 527

|||||

52 GTCTCTGCTACTGACTGTGTGTAGGGGACCGTTTGGCAGCTGTGCCGTC 101

|||||

527 iStyRGlyAspGlnValAspArgValLeuGlyThrLeuAlaAlaValPhe 543

```
|||||
102 ATTACGAGACACAGGTGGACAGGCTCTGGGACACCTGGCTGCTGCTTT 151
544 ValSerHisLeuHisAlaAspHisHisThrGlyLeuProSerIleLeuLe 560
|||||
152 GTGTCCACCTGCACGCAGATCACACACGGGCTTGCCAAAGTATCTTGCT 201
560 uGlnArgGluArgAlaLeuAlaSerLeuGlyLysProLeuHisProLeuL 577
|||||
202 GCAGAGAGAACGGCGCTTGCCATCTTTGGAAAGCGGCTTCACCCCTTGC 251
577 euValValAlaProAsnGlnLeuLysAlaTrpLeuGlnGlnTyrHisAsn 593
|||||
252 TGGTGGTTGCCCCCAACACAGCTCAAGCCCTGGCTCCAGCAGTACCAAC 301
594 GlnCysGlnGluValLeuHisHisIleSerMetIleProAlaLysCyste 610
|||||
302 CAGTGCACAGAGGTCTGCACACATCATGATGATTCCTGCCAAATGCCT 351
610 uGlnGluGlyAlaGluIleSerSerProAlaValGluArgLeuIleSerS 627
|||||
352 TCAGGAAGGGGTGAGATCTCCAGTCTCGAGTGGAAAGATTGATCAGTT 401
627 erLeuLeuArgThrCysAspLeuGluPheGlnThrCysLeuValArg 643
|||||
402 CCGTGTGGCAACATGTGATG. GAAGAGTTTCAGACCTGTCTGGTGCAG 450
644 HisCysLysHisAlaPheGlyCysAlaLeuValHisThrSerGlyTrpL 660
|||||
451 CACTGCAAGCATGCGTTTGGCTGTGGCTGGTGCACACCTCTGGCTGGAA 500
660 sValValTyrSerGlyAspThrMetProCysGluAlaLeuValArgMetG 677
|||||
501 AGTGTGCTATTCCGGGGACACCATGCGCTCGAGGCTCTGTCCGCGATGG 550
677 lYlYAspAlaThrLeuLeuIleHisGluAlaThrLeuGluAspGlyLeu 693
|||||
551 GGAAGATGCCACCTCTCGTATACATGAAGCCACCTTGAAGATGTTTG 600
694 GluGluGluAlaValGluLysThrHisSerThrThrSerGlnAlaIleSe 710
|||||
601 GAAGAGGAAGCAGTGGAAAGACACACAGCACACAGTCCCAAGCCATCAG 650
710 rValGlyMetArgMetAsnAlaGluPheIleMetLeuAsnHisPheSerG 727
|||||
651 CGTGGGATGCGGATGAACGCGAGTTTATTATGCTGAACCACTTCAGC. 699
727 lNArgTyrAlaLysValProLeuPheSerProAsnPheSerGluLysVal 743
|||||
700 ..AGCTATGCCAAGGTCCCTCTTCAGCCCACTTCAGCCGAGAAAGTG 747
744 GlyValAla 746
|||||
748 GGAGTTGCT 756
```

seq_name: gb_est69:BE260495

```
seq_documentation_block:
LOCUS BE260495 676 bp mRNA EST 13-JUL-2000
DEFINITION 601150702F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:3503184 5',
mRNA sequence.
ACCESSION BE260495
VERSION BE260495.1 GI:9131807
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1. (bases 1 to 676)
NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
```

Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: AFCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: L1CMI76 row: d column: 01
High quality sequence stop: 672.

FEATURES

Location/Qualifiers
source
1..676
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3503184"
/clone_lib="NIH_MGC_19"
/tissue_type="neuroblastoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
BASE COUNT 154 a 207 c 176 g 139 t
ORIGIN

alignment_scores:

Quality: 1180.00 Length: 225
Ratio: 5.244 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-434-382-2 x BE260495 ..

Align seg 1/1 to: BE260495 from: 1 to: 676

```
365 ThrGlnHisLeuValLeuAsnGluAsnCysAlaSerValHisAsnLeuAr 381
|||||
1 ACCCAGCACTTGGTCTGTAATGAGAACTGTGCTCAGTTTCACAACCTCG 50
381 gSerHisLysIleGlnThrGlnLeuAsnLeuIleHisProAspIlePheP 398
|||||
51 CAGCCACAAGATTCAACCCAGCTCAACCTCATCCACCGGACATCTTCC 100
398 roLeuLeuThrSerPheArgCysLysLysGluGlyProThrLeuSerVal 414
|||||
101 CCTGCTCACCAGTTTCCGCTGTAAGAGAGGGGCGCCACCCCTCAGTGTG 150
415 ProMetValGlnGlyCysLeuLeuLysTyrGlnLeuArgProArgAr 431
|||||
151 CCCATGTTTCAGGGTCAATGCTCTCAAGTACCAGCTCCGTCGCCAGGAG 200
431 gGluTrpGlnArgAspAlaIleIleThrCysAsnProGluGluPheIleV 448
|||||
201 GGAGTGGCAGAGGGATGCCATTATTACITGCAATCTCGAGGAATTCATAG 250
448 aLGlAlaLeuGlnLeuProAsnPheGlnGlnSerValGlnGluTyrArg 464
|||||
251 TTGAGGCGCTGCAGCTTCCCAACTTCCAGCAGAGCGTGCAGGAGTACAGG 300
465 ArgSerAlaGlnAspGlyProAlaProAlaGluLysArgSerGlnTyrPr 481
|||||
301 AGGATGGCAGCAGCGGCCACGCCAGCAGAGAGAAAGATCAGTACCC 350
481 oGluIleIlePheLeuGlyThrGlySerAlaIleProMetLysIleArgA 498
|||||
351 AGAAATCATCTCTCTTGGAAACAGGGTCTGCCATCCGATGAAGATTTCGAA 400
498 snValSerAlaThrLeuValAsnIleSerProAspThrSerLeuLeuLeu 514
```

```

|||||
401 ATGTCAAGTCCACACATGTCACATAGCCCGACACGCTCTCTGCTACGTG 450
515 AspCysGlyGluGlyThrPheGlyGlnLeuCysArgHisTyrGlyAspG1 531
|||||
451 GACTGTGGTGAGGCACATTTGGGACGCTGCGCGTCATTACGAGACCA 500
531 nValAspArgValLeuGlyThrLeuAlaLaValPheValSerHisLeuH 548
|||||
501 GTGGACAGGGTCTGGGACACCTGGCTGCTGCTGTTGTGTCCACCTGC 550
548 isAlaAspHisHisThrGlyLeuProSerIleLeuLeuGlnArgGluArg 564
|||||
551 ACCGATCACCACAGCGGCTTCCCAAGTATCTTGTGTCAGAGAGACGC 600
565 AlaLeuAlaSerLeuGlyLysProLeuHisProLeuValValAlaPr 581
|||||
601 GCCTTGGCATCTTTGGGAAGCGCTTACCCTTTGCTGGTGTGCCCC 650
581 oAsnGlnLeuLysAlaTrpLeuGln 589
|||||
651 CAACCAAGCTCAAGCGCTGGCTCCAG 675

```

seq_name: gb_est75:BE744197

seq_documentation_block:

LOCUS BE744197 949 bp mRNA EST 15-SEP-2000
 DEFINITION 601577168F1 NIH_MGC_9 Homo sapiens cDNA clone IMAGE:3838121 5',
 mRNA sequence.

ACCESSION BE744197

VERSION BE744197.1 GI:10158189

KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 949)

NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert_Strausberg@nih.gov

Tissue Procurement: DCTD/bnp

cDNA Library Preparation: Ling Hong/Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov

Plate: LLC524 row: g column: 18

High quality sequence stop: 668.

Location/Qualifiers

FEATURES

source

```

1..949
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3838121"
/clone_lib="NIH_MGC_9"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: ovary; Vector: pORF7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
```

BASE COUNT

ORIGIN

225 a 263 c 281 g 179 t 1 others

alignment_scores:

Quality: 1165.00

Length: 278

```

Ratio: 4.447 Gaps: 6
Percent Similarity: 94.245 Percent Identity: 90.647
alignment_block:
US:09-434-382-2 x BE744197
Align seg 1/1 to: BE744197 from: 1 to: 949
225 ProHisGlyValSerGlnArgArgGlyValArgAspSerSerLeuValVa 241
|||||
2 CCACATGGTGTAGCCAGAGAAGAGGGGTGAGGACTCTTCCCTGGTCGT 51
|||||
241 lAlaPheIleCysLysLeuHisLeuLysArgGlyAsnPheLeuValLeu 258
|||||
52 AGCTTTTCATCTGTAAGCTTCACTTAAAGAGAGGAAACTTCTTGGTGTCA 101
|||||
258 ysAlaLysGluMetGlyLeuProValGlyThrAlaAlaIleAlaProIle 274
|||||
102 AAGCAAGAGGAGATGGGCTCCCAAGTTGGGACAGTCCCATCGCTCCCATC 151
|||||
275 lIleAlaAlaValLysAspGlyLysSerIleThrHisGluGlyArgGlu 291
|||||
152 ATTGCTGCTCTCAAGGACGGGAAAGCATCACTCATGAAGAAGAGAGAT 200
|||||
291 eLeuAlaGluLeuCysThrProProAspProGlyAlaAlaPheValV 308
|||||
201 TTGGCTGAAGAGTGTGTACTCTCTCCAGATCTCTGGTGTGTC. TTGTGG 249
|||||
308 alValGluCysProAspGluSerPheIleGlnProIleCysGluAsnAla 324
|||||
250 TGGTAGAATGTCCAGATGAAGCTTCAATCAACCCATCTGTGAGAATGCC 299
|||||
325 ThrPheGlnArgTyrGlnGlyLysAlaAspAlaProValAlaLeuValVa 341
|||||
300 ACCTTTCAAGAGTACCAGGAAGGACAGATGCCCGCTGGCTGGTGGT 349
|||||
341 lHisMetAlaProAlaSerValLeuValAspSerArgTyrGlnGlnTrpM 358
|||||
350 TCACATGGCCCCAGCATCTGTGCTTGTGGACAGCAGGTACCAGCAGTGA 399
|||||
358 etGluArgPheGlyProAspThrGlnHisLeuValLeuAsnGluAsnCys 374
|||||
400 TGGAGAGGTTTGGGCTTGACCCAGCAGCATTTGGTCTCAATGAGAAGTGT 449
|||||
375 AlaserValHisAsnLeuArgSerHisLysIleGlnThrGlnLeuAsnLe 391
|||||
450 GCCTCAGTTCCAAACCTTCGACCCACACAGATTCAAACCCAGCTCAACT 499
|||||
391 uIleHisProAspIlePheProLeuLeuThrSerPheArgCysLysLys 408
|||||
500 CATCCACCCGGACATCTTCCCTGCTCACAGTTTCCGCTGTAAGAAGG 549
|||||
408 luGlyProThrLeuSerValProMetValGlnGlyGluCysLeuLeuLys 424
|||||
550 AGGGCCCCACCTCAGTGTGCCATTTGAGGTTCAGGGGTGAATGCCCTCAAG 599
|||||
425 TyrGlnLeuArgProArgArgGluTrpGlnArgAspAlaIleIleThr.C 441
|||||
600 TACCAGTCCGCTCCAGG. AGGGAGTGGCAGAGGATGCCATTATTACTTT 648
|||||
441 ysAsnProGluGluPheIleVal...GluAlaLeuGln...LeuProAsn 455
|||||
649 GCAATCTCTGAGGAATTCATAGCTTGAGGGCGGCTGCAAGGGCTTCCCAAC 698
|||||
456 PheGlnGln.SerValGlnGluTyrArgArgSerAla...GlnAspGlyP 471
|||||
699 CTCACGAGCAGGCGTGCAGGAGTATCCCGAGGAGTCCCGCAGGACGGCC 748
|||||
471 roAlaProAla...GluLysArgSerGlnTyrProGluIleIlePheLeu 486
|||||
749 CAGCCCCAGTCAGAGAAAGAAAGTAGTCAGGCACCAAGAAATCATTCTTCCT 798
|||||
487 GlyThrGlySerAlaIleProMetLysIle 496

```

```

||||||| ||| |||||||||
799 GGAACAGCTGTGGCACCCCGCATGAAGATC 828

seq_name: gb_est73:BE619259

seq_documentation_block:
LOCUS BE619259 823 bp mRNA EST 24-AUG-2000
DEFINITION 601473130F1 NTH_MGC_68 Homo sapiens cDNA clone IMAGE:3876223 5',
mRNA sequence.
ACCESSION BE619259
VERSION BE619259
KEYWORDS EST.
SOURCE BE619259.1 GI:9890197
human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: DCTD/DTP/Gazdar
cDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LCM623 row: k column: 08
High quality sequence stop: 695.
Location/Qualifiers
1. 823
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3876223"
/clone_lib="NIH_MGC_68"
/tissue_type="large cell carcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lung; Vector: pCMV-SPORT6; Site.1: NotI;
Site.2: SalI; Cloned unidirectionally. Primer: Oligo dr.
Average insert size 1.8 kb. Library constructed by Life
Technologies."
BASE COUNT 185 a 219 c 259 g 160 t
ORIGIN

alignment_scores:
Quality: 1164.50 Length: 284
Ratio: 4.603 Caps: 3
Percent Similarity: 89.085 Percent Identity: 84.859

alignment_block:
US-09-434-382-2 x BE619259 ..
Align seg.1/1 to: BE619259 from: 1 to: 823
519 G|ThrPheGlyGlnLeuCysArgHisTyrGlyAspGlnValAspArgVa 535
|||||
3 GGGACAGTTGGGCAGCTGTGCCTCATTTACGGAGACCGAGGTGGACAGGT 52
535 lLeuGlyThrLeuAlaAlaValPheValSerHisLeuHisAlaAspHisH 552
|||||
53 CCTGGGCACCTTGGCTACTGTGT. GTGTCCCACTGCACGCAGATCACC 101
552 isthrGlyLeuProSerIleLeuLeuGlnArgGluArgAlaLeuAlaSer 568
102 ACACGGGCTTGGCAGATATCTTCTGCTCAGAGAGACGGCGCTTGGCATCT 151
569 LeuClyLysProLeuHisProLeuLeuValAlaProAsnGlnLeuLys 585
152 TTGGGAAGCCGCTTCACCTTTCGTGGTGGTTCGCCCAACCACTCA 201

```

JOURNAL
COMMENT

Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: DCTD/BTP
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LLCM522 row: c column: 06
High quality sequence stop: 726.
Location/Qualifiers

FEATURES

source

1. 984
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3837533"
/clone_lib="NIH_MGC_9"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: ovary; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCAGCAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 227 a 259 c 316 g 182 t
ORIGIN

alignment_scores:

Quality: 1158.50 Length: 253
Ratio: 4.634 Gaps: 2
Percent Similarity: 98.814 Percent Identity: 96.838

alignment_block:

US-09-434-382-2 x BE744876

Align seg 1/1 to: BE744876 from: 1 to: 984

225 ProHisGly.ValSerGlnArgGlyValArgAspSerLeuValVal 241
|||||
2 CCACATGGTGGTTAGCAGAGAGAGGGGCTCAGGACTCTTCCCTGGTGC 51
241 aLaPheIleCysLysLeuHisLeuLysArgGlyAsnPhelLeuVal 257
|||||
52 TAGCTTTCATCTGTAGCTTCACCTTAAGAGAGAGAACTCTTGGTGCTC 101
258 LysAlaLysGluMetGlyLeuProValGlyThrAlaAlaIleAlaProTl 274
102 AAAGCAAGGAGATGGGCTCCAGTTGGACAGCTGCCATGCCATCCAT 151
274 eileAlaValLysAspGlyLysSerIleThrHisGluGlyArgGluI 291
152 CATTCGTGCTGTCAAGGAGCGGAAAGAGCATCACTCATGAAGGAAGAGA 200
291 leLeuAlaGluGluLeuCysThrProProAspProGlyAlaAlaPheVal 307
201 T.TTGGCTGAAGAGCTGTACTCTCCAGATCCTGGTGCTGC.TTGTGG 249
308 ValValGluCysProAspGluSerPheIleGlnProIleCysGluAsnAl 324
250 GTGGTAGATGTCACAGATCAAGCTTCATTCACCCATCTGTGAGATGC 299
324 aThrPheGlnArgTyTGlnGlyLysAlaAspAlaProValAlaLeuVal 341
300 CACCTTTCAGAGGTACCAAGGAAGGAGATGCCCGCTGGCTGGTGG 349
341 aHisMetAlaProAlaSerValLeuValAspSerArgTyTGlnGlnTrp 357
350 TTCATATGGCCCCAGCATCTGTGCTTGTGGACAGCAGGTACACGAGTGG 399

358 MetGluArgPheGlyProAspThrGlnHisLeuValLeuAsnGluAsnCy 374
|||||
400 ATGGAGAGGTT.GGGCTGTACACCCAGCACTTGTCTCTGAATGAGAACTG 448
374 sAlaSerValHisAsnLeuArgSerHisLysIleGlnThrGlnLeuAsnL 391
|||||
449 TGCTCTCAGTTCAACAGCTTGGCAGCCACAGATTCAAAACCCAGCTCAAC 498
391 eulIleHisProAspIlePheProLeuLeuThrSerPheArgCysLysLys 407
|||||
499 TCATCCACCGGACATCTTCCCTCTGCTCACCAGTTTCCGCTGTAGAAG 548
408 GluGlyProThrLeuSerValProMetValGlnGlyGluCysLeuLeuLy 424
|||||
549 GAGGCGCCACCCCTCAGTGTGCCCATGTTTCAGGGTCAATGGCTCCTCAA 598
424 sTyrglnLeuArgProArgArgGluTrpGlnArgAspAlaIleIleThrC 441
599 GTACCAAGCTCCGTCAGGAGGAGTGGCAGAGGGATGCCATTATTACTT 648
441 ysAsnProGluGluPheIleValGluAlaLeuGlnLeuProAsnPhelGln 457
|||||
649 GCATCTCTGAGGAATTCATAGTTGAGGCGGTGAG.CTTCCCAACTTCAG 697
458 GlnSerValGlnGluTyArgArgSerAlaGlnAspGlyProAlaProAl 474
|||||
698 CACAGCGTGCAGGAGTACAGGAGGAGTGCAGAGGCC...AGCCGAGC 744
474 aGluLys 476
|||||
745 AGAGAAA 751

seq_name: gb_est76:BE867512

seq_documentation_block:

LOCUS BE867512 812 bp mRNA EST. 27-SEP-2000
DEFINITION 601443010F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3847226 5',
mRNA sequence.

ACCESSION BE867512

VERSION BE867512.1 GI:10316288

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 812)

NIH-MGC <http://www.ncbi.nlm.nih.gov/MGC/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert_Strausberg@nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLCM548 row: c column: 03

High quality sequence stop: 686.

FEATURES

source

1. 812
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3847226"
/clone_lib="NIH_MGC_65"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.8 kb. Library constructed by Life

BASE COUNT 200 a 216 c 220 g 176 t
 ORIGIN
 alignment_scores:
 Quality: 1154.50 Length: 270
 Ratio: 4.655 Gaps: 5
 Percent Similarity: 91.852 Percent Identity: 90.000
 alignment_block:
 US-09-434-382-2 x BE867512 ..

Align seg 1/1 to: BE867512 from: 1 to: 812

```

242 AlaPheIleCysLysLeuHisLeuLysArgGlyAsnPheLeuValLeuLys 258
|||||
1 GCTTTCATCTGTAAGCTTCACTTAAAGAGAGAAACTTCTGGTGCTCA 50
258 sAlaLysGluMetGlyLeuProValGlyThrAlaAlaIleAlaProIleI 275
|||||
51 AGCAAAGAGATGGGCCCTCCAGTTGGGACAGCTGCCATCGTCCCATCA 100
275 leAlaAlaValLysAspGlyLysSerIleThrHisGluGlyArgGluIle 291
|||||
101 TTGCTGCTCAAGACCGGAAAGACATCACTCAAGAGGAGAGAGATT 150
292 LeuAlaGluGluLeuCysThrProProAspProGlyAlaAlaPheValVa 308
|||||
151 TTGGCTGAAGAGCTGTGTACTCTCCACAGATCCTGGTGCTGC.TTGTGT 199
308 lvaIgluCysProaspGluSerPheIleGlnProIleCysGluAsnAlaT 325
200 GGTAGAATGTCAGATGAAGCTTCATTCAACCCCATCTGTGAGAAATGCCA 249
325 hrPheGlnArgTyrGlnGlyLysAlaAspAlaProValAlaLeuValVal 341
|||||
250 CCTTTCAGAGTACCAAGAAAGGAGATGCCCGTGGCTTGTGTGT 299
342 HisMetAlaProAlaSerValLeuValAspSerArgTyrGlnGlnTrpMe 358
300 CACATGGCCCCAGCATCTGTCTGTGGACAGCAGGTACCGAGCAGTGAT 349
358 tGluArgPheGlyProAspThrGlnHisLeuValLeuAsnGluAsnCysA 375
350 GGAGAGGTTTGGGCTGACACCCAGCACTTGGTCTGTAATGAGAACTGTG 399
375 laSerValHisAsnLeuArgSerHisLysIleGlnThrGlnLeuAsnLeu 391
|||||
400 CCTAGTTTCAACACTTCGACGCCACAAAGATTCAACCCAGCTCAACCTC 449
392 IleHisProAspIlePheProLeuLeuThrSerPheArgCysLysLysG1 408
450 ATCCACCCGGACATCTTCCCTCTGCTCACCAGTTTCCGCTGTAAGAAGGA 499
408 uGlyProThrLeuSerValProMetValGlnGlyCysLeuLeuLysT 425
500 GGGCCCCACCTCAGTGTGCCATGGTTCAGGGTGAATGCCCTCTCAAGT 549
425 yrGlnLeuArgProArgArg.GluTrp.GlnArgAspAlaIleIleThr. 440
550 ACCAGCTCCGTCCAGGAGGCGGTGGGACAGGAGATGCCATTATTACTC 599
441 CysAsnProGluGluPheIleValIcIuAlaLeuGlnLeuProAsnPheG1 457
600 TGCAATCTGAGGAATTCATAGTTGAGGCGCTGCAGCTTCCCACTTCAG 648
457 nGlnSerValGlnGluTyrArgArgSerAlaGlnAspGlyProAlaProA 474
649 .CAGAGCGGTGCAGGATACAGGAGAGCTGCCGCCAGGCGGCCAGCCAG 698
474 laGluLysArgSerGlnTyrProGluIleIlePheLeuGlyThrGlySer 490
|||||

```

```

699 CAGAGAAAAGAGTCTAGTACCAGGAATCATC.....TCCTGGGACGGT 742
491 AlalleProMetLysIle.....ArgAs 498
|||||
743 CTGCATCCGATGAAGATCGAATGTCGGGCCACTGTTAATATTAGCCGAAA 792
498 nVal 499
|||||
793 CGTT 796

```

seq_name: gb_est75:BE747163

seq_documentation_block:

LOCUS BE747163 992 bp mRNA EST 15-SEP-2000
 DEFINITION 601577254F1 NIH_MGC_9 Homo sapiens cDNA clone IMAGE:3838237 5',
 mRNA sequence.

ACCESSION BE747163

VERSION BE747163.1 GI:10161155

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 992)

REFERENCE NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.

AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert_Strausberg@nih.gov

Tissue Procurement: DCTD/DTP

cDNA Library Preparation: Ling Hong/Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov

Plate: LCM524 row: 1 column: 14

High quality sequence stop: 781.

FEATURES

source

```

1..992
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3838237"
/clone_lib="NIH_MGC_9"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: ovary; Vector: pOT87; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGACAGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
```

BASE COUNT 225 a 287 c 313 g 167 t

ORIGIN

alignment_scores:

Quality: 1140.00 Length: 275

Ratio: 4.488 Gaps: 5

Percent Similarity: 92.364 Percent Identity: 89.818

alignment_block:

US-09-434-382-2 x BE747163 ..

Align seg 1/1 to: BE747163 from: 1 to: 992

331 GlyLysAlaAspAlaProValAlaLeuValValHisMetAlaProAlaSe 347

1 GGAAAGCAGATGCCCGCTGGC.CTGGTGGTTACATGGCCCCCAGCATC 49

347 rValLeuValAspSerArgTyrGlnGlnTrpMetGluArgPheGlyProA 364


```

|||||
50  TGTGCTTGTGGACAGCAGGTACACAGCTGGATGGAGGTTT.GGCGCTG 98
|||||
364 spThrGlnHisLeuValLeuAsnGluAsnCysAlaSerValHisAsnLeu 380
|||||
99  ACACCCAGCAGCTTGGTCTCTGAATGAGAAGTGTGCTCAGTTCAACACCTT 148
|||||
381 ArgSerHisIleGlnThrGlnLeuAsnLeuIleHisProAspIlePh 397
|||||
149 CGCAGCCACAAGATTCAAAACCCAGCTCAACCTCATCCACCGGACATCTT 198
|||||
397 eProLeuLeuThrSerPheArgCysIleGluGlyProThrLeuSerV 414
|||||
199 CCCCCCTGCTCACCGATTTCCGCTGTAGAGAGGAGGCCGCCACCTCAGTG 248
|||||
414 alProMetValGlnGlnGlyCysLeuLeuLysTyrGlnLeuArgProArg 430
|||||
249 TGCCCATGGTTTCAGGTGAATGCTCTCAAGTACCAGCTCCGTCGCCAGG 298
|||||
431 ArgGluTrpGlnArgAspAlaIleIleThrCysAsnProGluLupheII 447
|||||
299 AGGAGTGGCAGAGGATGCCATTATTTACTTGAATCCTGAGGAATTCAT 348
|||||
447 eValGluAlaLeuGlnLeuProAsnPheGlnGlnSerValGlnGluTyrA 464
|||||
349 AGTTGAGGCGCTCCAGCTTCCCAACCTTCCAGCAGAGCGTGCGAGGATACA 398
|||||
464 rgArgSerAlaGlnAspGlyProAlaProAlaGluLysArgSerGlnTyr 480
|||||
399 GGAGGAGTGGCAGGAGCGG.CCAGCCCCAGCAGCAGAGAAAGAACTCAGTAC 447
|||||
481 ProGluIleIlePheLeuGlyThrGlySerAlaIleProMetIleAr 497
|||||
448 CCAGAAATCATCTCTCTTGAACAGGCTGTGCCATCCGATGAAGATTTCG 497
|||||
497 gAsnValSerAlaThrLeuValAsnIleSerProAspThrSerLeuLeu 514
|||||
498 AATGTGTCAGTGCCACACTTGTCAACATAAGCCCCGACACGCTTTGCTAC 547
|||||
514 euAspCysGlyGluGlyThrPheGlyGlnLeuCysArgHisTyrGlyAsp 530
|||||
548 TGGACTGTGTGAGGCGACATTTGGCGAGCTGTGCCCTCATTCAGGAGAC 597
|||||
531 GlnValAspArgValLeuGlyThrLeuAlaAlaValPheValSerHisLe 547
|||||
598 CAGGTGGACAGGCTCTGGGACACCTGGCTGTGTGTTGTGTCCACCT 647
|||||
547 uHisAlaAspHisThrGlyLeuProSerIleLeuLeuGlnArgGluA 564
|||||
648 GCACGCGAGAT...CACACAGGCGCTGCAAGTATCTTGTCTGCAGAGAGAAC 694
|||||
564 rgAlaLeuAlaSerLeuGlyLysProLeuHisProLeuLeuValValAla 580
|||||
695 GCCTGCTG...GATCTGGGAAAGCGGTTCACCTTAGTGG...TGGTGC 737
|||||
581 ProAsnGlnLeuLysAlaTrpLeuGlnGlnTyrHisAsnGlnCysGlnI 597
|||||
738 CCCAACCCAGTCAAAAGGGG.....TCAGCAGTACACAAACCAAGTGCAGGAG 781
|||||
597 uValLeuHisHisIleSerMetile 605
|||||
782 GTCTGCC...CACATAGGATGATC 803

```

seq_name: gb_est75:BE794311

```

seq_documentation_block: 688 bp mRNA EST 20-SEP-2000
LOCUS BE794311
DEFINITION 601591442F1 NIH_MGC_7 Homo sapiens cDNA clone IMAGE:3945538 5',
mRNA sequence.
ACCESSION BE794311
VERSION BE794311.1 GI:10215496
KEYWORDS EST.
SOURCE human.

```

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 688)

AUTHORS

NIH-MGC <http://www.ncbi.nlm.nih.gov/MGC/>.

TITLE

National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL

Unpublished (1999)

COMMENT

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert_Strausberg@nih.gov

Tissue procurement: DCTD/DTp

cDNA Library Preparation: Ling Hong/Rubin Laboratory

cDNA Library Arrayed by: the I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov

Plate: LLCM804 row: c column: 11

High quality sequence stop: 688.

FEATURES

source

Location/Qualifiers

1..688

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:3945538"

/clone_lib="NIH_MGC_7"

/tissue_type="small cell carcinoma"

/cell_line="MGC3"

/lab_host="DH10B (phage-resistant)"

/note="Organ: lung; Vector: pOTB7; Site_1: XhoI; site_2:

EcoRI; cDNA made by oligo-dT priming. Directionally

cloned into EcoRI/XhoI sites using the following 5'

adaptor: GGCACAG(G). Size-selected >500bp for average

insert size 1.8kb. Library constructed by Ling Hong in

the laboratory of Gerald M. Rubin (University of

California, Berkeley) using ZAP-cDNA synthesis kit

(Stratagene) and Superscript II RT (Life Technologies)."

169 a 190 c 185 g 144 t

BASE COUNT

ORIGIN

alignment_scores:

Quality: 1131.00 Length: 230

Ratio: 4.982 Gaps: 1

Percent Similarity: 98.696 Percent Identity: 98.696

alignment_block:

US-09-434-382-2 x BE794311

Align seg 1/1 to: BE794311 from: 1 to: 688

258 LysAlaLysGluMetGlyLeuProValGlyThrAlaAlaIleAlaProII 274

|||||

3 AAAGCAAAGGAGATGGGCTCCAGTTGGGACAGCTGCCATCGCTCCAT 52

|||||

274 eIleAlaAlaValLysAspGlyLysSerIleThrHisGluGlyArgGluI 291

|||||

53 CATTTGCTGCTGTCAAGGACGGGAAAGCATCATCATCAAGAAAGAGAGA 102

|||||

291 leLeuAlaGluLeuLeuCysThrProProAspProGlyAlaAlaPheVal 307

|||||

103 TTTTGGCTGAAGAGCTGTACTCTCCAGATCTCTGGTGTCTTTGTG 152

|||||

308 ValValGluCysProAspGluSerPheIleGlnProIleCysGluAsnAl 324

|||||

153 GTGTAGAATGTCCAGATGAAGCTTCAATCAACCCATCTGTGAGATGC 202

|||||

324 aThrPheGlnArgTyrGlnGlyLysAlaAspAlaProValAlaLeuValV 341

|||||

203 CACCTTTTCAGAGGTACCAAGGAGGAGAGGAGAGTGGCTTGGTGG 252

|||||

341 alHisMetAlaProAlaSerValLeuValAspSerArgTyrGlnGlnTrp 357

|||||

252 TTTTGTGGGCCCCAGCATCTGTGCTTTGTGGACAGGATACCGACAGTGG 302

|||||

358 MetGluArgPheGlyProAspThrGlnHisLeuValLeuAsnGluAsnCy 374
 303 ATGGAGAGTTTGGGCTGACACCCAGCAGCTGGTCTGTAATGAGAATG 352

374 sAlaSerValHisAsnLeuArgSerHisLysIleGlnThrGlnLeuAsnL 391
 353 TGCCCTCAGTTCAACCTTCGAGGACCAAGATTCAAACCCAGCTCAACC 402

391 euIleHisProAspIlePheProLeuLeuThrSerPheArgCysLysLys 407
 403 TCATCCACCGGACATCTCCCTGCTCACCAGTTTCGGTGTAGAAG 452

408 GluGlyProThrLeuSerValProMetValGlnGlyGluCysLeuLeuL 424
 453 GAGGGCCCACTC. AGTGTGCCATGTTTCAGGTGAATGCTCTCAA 501

424 sTyGlnLeuArgProArgGluTrpGlnArgAspAlaIleIleThrC 441
 502 GTACCAAGTCCGTCGAGGAGGAGTGCGAGAGGATGCCATTATTACTT 551

441 yAsnProGluGluPheIleValGluAlaLeuGlnLeuProAsnPheGln 457
 552 GCAATCCTGAGGAATTCATAGTTGAGGCGCTGCAGCTTCCCAACTTCCAG 601

458 GlnSerValGlnGluTyArgArgSerAlaGlnAspGlyProAlaProAl 474
 602 CAGAGCGTGCAAGGAGTACAGGAGGAGTGCGCAGGACGG. CCAGCCCCAGC 650

474 aCluLysArgSerGln.TyrProGluIleIlePheLeu 486
 651 AGAGACAAGAGTCAAGTACCCAGAAATCATCTTCTCTG 688

seq_name: gb_est77:BE892893

seq_documentation_block: 790 bp mRNA EST 29-SEP-2000

LOCUS BE892893 601435738F1 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:3920792 5',

DEFINITION mRNA sequence.

ACCESSION BE892893

VERSION BE892893.1 GI:10353525

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AUTHORS 1 (bases 1 to 790)

TITLE NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.

JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)

COMMENT Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

Tissue Procurement: ATCC/DCTD/DRP

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

... Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLCW739 row: 1 column: 09

High quality sequence stop: 662.

FEATURES

source

BASE COUNT 185 a 221 c 220 g 164 t
 ORIGIN

alignment_scores:

Quality: 1123.00 Length: 267

Ratio: 4.528 Gaps: 10

Percent Similarity: 92.884 Percent Identity: 88.764

alignment_block:

US-09-434-382-2 x BE892893

Align seg 1/1 to: BE892893 from: 1 to: 790

302 ProGlyAlaAlaPheValValGluCysProAspGluSerPheIleGl 318

2 CTGTGTCTGCTTTTGTGTGTAGATGTCAGATGAAGCTTCATTCA 51

318 nProIleCysGluAsnAlaThrPheGlnArgTyrGlnGlyLysAlaAspA 335

52 ACCCATCTGTGAGATGCCACCTTTCAGAGGTACCAAGGAAGGCAGATG 101

335 laProValAlaLeuValValHisMetAlaProAlaSerValLeuValAsp 351

102 CCCCCGTGGCCTTGTGTGTTCATATGCCCCAGCATCTGTCTGTGGAC 151

352 SerArgTyrGlnGlnTrpMetGluArgPheGlyProAspThrGlnHisLe 368

152 AGCAGGTACCCAGCAGTGGATGGAGAGTTTGGGCTGCACCCAGCAGCTT 201

368 uValLeuAsnGluAsnCysAlaSerValHisAsnLeuArgSerHisLysI 385

202 GGTCTCTGAATGAGAACTGTGCTCAGTTCACAACCTTCGCAGCCACAAGA 251

385 leGlnThrGlnLeuAsnLeuIleHisProAspIlePheProLeuLeuThr 401

252 TTCAAAACCCAGCTCAACCTCATCCACCGGACATCTTCCCTCTGCTCAC 301

402 SerPheArgCysLysLysGluGlyProThrLeuSerValProMet. ValG 418

302 AGTTTCCCTGTAAAGAGAGGGCCCCCCTCAGTGTGCCATGGGTTC 351

418 InGlyGluCysLeuLeuLysTyrGlnLeuArgProArgGluTrpGln 434

352 AGGGTGAATGCTCTCCTCAAGTACAGCTCCGTCGCCAGGAGGAGTGGCAG 401

435 ArgAspAlaIleIleThrCysAsnProGluGluPheIleValGluAlaLe 451

402 AGGGATGCGCATTTACTTGAATCCTGAGGAATTCATAGTTGAGCGCT 451

451 uGlnLeu. ProAsnPheGlnSerValGlnGluTyrArgArgSerAla 467

452 GCAGTTTCCCAACTTCCAGCAGCGTGCAGGAGTACAGGAGGAGTGGC 501

468 GlnAspGlyProAlaProAlaGluLysArgSerGlnTyr. ProGluIleI 484

502 CAGGACGCCGCCAGCCAGCAGAGAGAAAGATCAGTACCCAGAAATCA 551

484 lePheLeuGlyThrGlySerAlaIleProMet. LysIleArg. AsnVal 500

552 TCTTCTTGGAAACAGGGTCTGCCATCCCGGAGAAAGATTGCAACATGCA 601

500 erAlaThrLeu. ValAsnIleSerProAspThrSerLeuLeuLeuAsp 515

602 GTGCCACACTTGGTCTACATAAAGCCCCGACACGCTCTCTGCTACTGGAC 651

516 CysGlyGluGlyThrPhe. GlyGlnLeuCysArgHisTyrGly. AspGln 531

652 TGTGTGTAGGGCACATTGGGGCAAGTGTGGCGCCATTACGGAAGACCCAG 701

532 ValAspArgValLeuGlyThr. LeuAlaAlaValPheVal...SerHisL 547

702 GTGGCCAGGGTCTCTGGGGACACCTGGGCTTGTGGGTTTGGGTTTCCACC 751

547 euHisAlaAspHisThrGlyLeuProSerIle 558
 |||||
 752 GGCACGACAGATCACCAAGCGGCTTGACAAAT 786

seq_name: gb_est77:BE900936

seq_documentation_block: 745 bp mRNA EST 29-SEP-2000
 LOCUS BE900936 601674206F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:3957240 5',
 mRNA sequence.

ACCESSION BE900936
 VERSION BE900936.1 GI:10389609
 KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 745)
 NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert_Strausberg@nih.gov
 Tissue Procurement: ATCC

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
 Plate: LCM834 row: k column: 01
 High quality sequence stop: 734.
 Location/Qualifiers

FEATURES
 source

1..745
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3957240"
 /clone_lib="NIH_MGC_21"
 /tissue_type="choriocarcinoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: placenta; Vector: pOTB7; Site_1: XhoI;
 Site_2: EcoRI; cDNA made by oligo-dT priming.
 Directionally cloned into EcoRI/XhoI sites using the
 following 5' adaptor: GGCACGAG(G). Size-selected >500bp
 for average insert size 1.8kb. Library constructed by
 Ling Hong in the laboratory of Gerald M. Rubin (University
 of California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies)."
 BASE COUNT 165 a 206 c 215 g 159 t

alignment_scores:

Quality: 1120.00 Length: 252
 Ratio: 4.590 Gaps: 7
 Percent Similarity: 96.825 Percent Identity: 95.635

alignment_block:

US-09-434-382-2 x BE900936

Align seg 1/1 to: BE900936 from: 1 to: 745

509 AspThrSerLeuLeuLeuAspCysGlyGluGlyThrPheGlyGlnLeuCy 525
 |||||
 2 GACACGCTCTGCTACTGACTGTGGTGGGACACATTTGGGAGCTGTG 51

525 sArgHisTyrGlyAspGlnValAspArgValLeuGlyThrLeuAlaVal 542
 |||||
 52 CCCTCATTTACGGAGACAGGTGGACAGGGTCTGGGACCCCTGGCTGCTG 101

542 alpHeValSerHisLeuHisAlaAspHisThrGlyLeuProSerIle 558
 |||||

102 TGTT.GTGTCCCACTGCACGCGAGATCACACACAGGGCTTGCCAAGTATC 150
 559 LeuLeuGlnArgGluArgAlaLeuAlaSerLeuGlyLysProLeuHisPr 575
 |||||
 151 TTGCTGCAGAGACACGCGCTTGGCATCTTTGGAAAGCGCTTACCC 200
 575 oLeuLeuValValAlaProAsnGlnLeuLysAlaTrpLeuGlnGlnTyrH 592
 |||||
 201 TTTGCTGGTGGTTGGCCCAACAGCTCAAGGCTGGCTCCAGCAGTACC 250
 592 isAsnGlnCysGlnGluValLeuHisHisIleSerMetIleProAlaLys 608
 |||||
 251 ACAACACAGTGCAGAGGCTCTGCACCAACATCATGATGATTCCTGCCAA 300
 609 CysLeuGlnGluGlyAlaGluIleSerSerProAlaValGluArgLeuI 625
 |||||
 301 TGCCTTTCAGGAAGGGGTGAGATCTCCAGTCTCCTGCAAGTGAAGATTGAT 350
 625 eSerSerLeuLeuArgThrCysAspLeuGluGluPheGlnThrCysLeuV 642
 |||||
 351 CAGTTCCGCTGTTCCGAACATGTGATTTGGAAGAGTTTCAGACCTGTCTGG 400
 642 aArgHisCysLysHisAlaPheGlyCysAlaLeuValHisThrSerGly 658
 |||||
 401 TGCAGGCACTGCAAGCATGCTTTGGCTGTGCTGGTGCACACCTCTGGC 450
 659 TrpLysValValTyrSerGlyAspThrMetProCysGluAlaLeuValAr 675
 |||||
 451 TGGAAAGTGGTCTATTCCGG.GACACCATCCCTCGGAGGCTCTGGTCCG 499
 675 gMetGlyLysAspAlaThrLeuLeuIleHisGluAlaThr.LeuGluAsp 691
 |||||
 500 GATGGGGAAGATGCCACCTCCTGATACATGAAGCCACCCCTGGAAGAT 549
 692 GlyLeuGluGluGluAlaValGluLysThrHisSerThrThrSerGlnAl 708
 |||||
 550 GGTGTGAAGAGGAAGCAGTGCACCAAGACACACAGCACAACGTCCTCAAGC 599
 708 alLeSerValGlyMetArgMetAsn.AlaGluPhe.IleMetLeuAsnHi 724
 |||||
 600 CATCAGCTGGGATCGGGATGAACCGCGGAGTTCATTATGCTGAACCA 649
 724 sPheSerGlnArgTyrAlaLys..ValProLeuPheSerProAsnPheSe 740
 |||||
 650 ATTACAGCAAGGATATGCCAAAGGTCCTCTTTCAGCCCAAAATTTTCAG 699
 740 r.GluLysValGly.ValAlaPheAsp.HisMetLysValCys 753
 |||||
 700 CGGAGAAAGTGGGAGTGGGCTTTGACCCACATGAAGGTCTGC 742

seq_name: gb_est77:BE902696

seq_documentation_block:

LOCUS BE902696 735 bp mRNA EST 29-SEP-2000
 DEFINITION 601677393F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:3959926 5',
 mRNA sequence.
 ACCESSION BE902696
 VERSION BE902696.1 GI:10393148
 KEYWORDS EST.
 SOURCE human.

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 735)

AUTHORS

NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.

TITLE

National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL

Unpublished (1999)

COMMENT

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert_Strausberg@nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Ling Hong/Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov Plate: LLCM841 row: j column: 23
 High quality sequence stop: 732.
 Location/Qualifiers
 1. .735
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3959926"
 /clone_lib="NIH_MGC_21"
 /tissue_type="choriocarcinoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: placenta; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."
 BASE COUNT 168 a 206 c 203 g 157 t 1 others
 ORIGIN

alignment_scores:
 Quality: 1109.00 Length: 246
 Ratio: 4.660 Gaps: 2
 Percent Similarity: 96.748 Percent Identity: 95.528
 alignment_block:
 US-09-434-382-2 x BE902696

Align seg 1/1 to: BE902696 from: 1 to: 735

398 ProLeuLeuThrSerPheArgCysLysGluGlyProThrLeuSerVa 414
 3 CCCCTGCTCACCACTTCCCGCTGTAAAGAGGGCCACCCCTCAGTGT 52
 414 IProMetValGlnGlyCysLeuLeuLysTyrGlnLeuArgProArgA 431
 53 GCCCATGTTCCAGGATGAATGCTCCTCAAGTACCAGCTCCGTCCAGGA 102
 431 rgGluTrpGlnArgAspAlaIleIleThr.CysAsnProGluGluPheII 447
 103 GGGAGTGGCAGAGGATGCCATTATTTACTTNGCAATCTCGAGGAATTCA 152
 447 eValGluAlaLeuGlnLeuProAsnPheGlnGlnSerValGlnGluTyrA 464
 153 ACTTGAGCGCTGCAGCTTCCCACTCCAGCAGCGGTGAGGAGTACA 202
 464 rgArgSerAlaGlnAspGlyProAlaProAlaGluLysArgSerGlnTyr 480
 203 GGAGGAGTGGCAGCAGCGCCAGCCCGCAGCAGAGAAAGATCAGTAC 252
 481 ProGluIleIlePheLeuGlyThrGlySerAlaIleProMetLysIleAr 497
 253 CCAGAAATCATCTTCTTGGAAACAGGGTCTCCCATCCCGATGAAGATTTCG 302
 497 gAsnValSerAlaThrLeuValAsnIleSerProAspThrSerLeuLeuL 514
 303 AAATGTCAGTCCACACTTGTCAACATAAGCCCGCAGACAGTCTGTGTAC 352
 514 euAspCysGlyGluGlyThrPheGlyGlnLeuCysArgHisTyrGlyAsp 530
 353 TGGACTGTGGTGGAGGCACATTTGGGCAGCTGTGCCGTCAATTCGGAGAC 402
 531 GlnValAspArgValLeuGlyThrLeuAlaValPheValSerHisLe 547
 403 CAGGTGGACAGGGTCTGGGCACCTGGCTGTGT.TGTGCCACCT 451
 547 uHisAlaAspHisHisThrGlyLeuProSerIleLeuLeuGlnArgGluA 564

452 GCACCGCAGATCACCACACCGGCTTGCCAAAGTATCTTGTCTGCAGAGAAC 501
 564 rGAlaLeuAlaSerLeuGlyLysProLeuHisProLeuLeuValValAla 580
 502 GCGCCTTGGCATC.TTGGGAAAGCCGCTTCACCCITTTGCTGGTGGTGGC 550
 581 ProAsnGlnLeuLysAlaTrpLeuGlnGlnTyrHisAsnGlnCysGlnG 597
 551 CCCAACCCAGCTCAAAAGCCTGGCTCCAGCAGTAGTACCACACCCAGTGC 600
 597 uValLeuHisHisIleSerMetIleProAlaLysCysLeuGlnGluGlyA 614
 601 GGTCTGCAACCATCATGATGATTCCTGC.AAATGCCCTTCAGGAAGGG 649
 614 laGluIleSerProAlaValGluArgLeuIleSerSerLeuLeuArg 630
 650 CTGAGATCTCCAGTCTCGAGTGGAAAGATTGATCAGTTCGCTGTTCGA 699
 631 ThrCysAspLeu.GluGluPheGlnThrCysLeu 641
 700 ACATGTGATTTGGGAAGAGTTTCAGAACTGTCTG 733

seq_name: gb_est70:BE383336

seq_documentation_block:
 LOCUS BE383336 664 bp mRNA EST 21-JUL-2000
 DEFINITION 601298249F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:3628308 5', mRNA sequence.
 ACCESSION BE383336
 VERSION BE383336.1 GI:9328701
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (Bases 1 to 664)
 AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Tel.: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue procurement: ATCC
 cDNA Library Preparation: Ling Hong/Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov Plate: LLCM313 row: a column: 13
 High quality sequence stop: 662.

FEATURES

source
 1. .664
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3628308"
 /clone_lib="NIH_MGC_19"
 /tissue_type="neuroblastoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."
 Note: This is a NIH_MGC Library.
 BASE COUNT 154 a 199 c 174 g 137 t
 ORIGIN

alignment_scores:
 Quality: 1086.50 Length: 221
 Ratio: 5.053 Gaps: 1

Percent Similarity: 97.285 Percent Identity: 95.023

alignment_block:

US-09-434-382-2 x BE383336 ..

Align seg 1/1 to: BE383336 from: 1 to: 664

```
365 ThrGlnHisLeuValLeuAsnGluAsnCysAlaSerValHisAsnLeuAr 381
1 ACCCAGCACTGGTCTCTGAATGAGAACTGGCTCAGTTCAACAACCTGC 50
381 gSerHisLysIleGlnThrGlnLeuAsnLeuLysHisProAspIlePhep 398
51 CAGCCACAGATTCAACCCAGCTCAACCTCATCCACCGGAGATCTTC 100
398 roLeuLeuThrSerPheArgCysLysLysGluGlyProThrLeuSerVal 414
101 CCTGCTCACCAGTTTCCTGCTGTGAAGAGGAGGCCCCACCTCAGTGTG 150
415 ProMetValGlnGlyCysLeuLeuLysTyrGlnLeuArgProArgAr 431
151 CCCATGGTTTCAGGGTGAATTCCTCTCAAGTACCAAGTACCGTCCCAAGGAG 200
431 gGluTrpGlnArgAspAlaIleThrCysAsnProGluGluPheIleV 448
201 GGAGTGGCAGAGGATGCCATTATTACTTGCATCTCAGGAATTCATAG 250
448 aGluAlaLeuGlnLeuProAsnPheGlnSerValGlnGluTyrArg 464
251 TTGAGCGCTGCAGCTTCCCACTCCAGCAGCGTGCAGGAGTACAGG 300
465 ArgSerAlaGlnAspGlyProAlaProAlaGluLysArgSerGlnTyrPr 481
301 AGGAGTGCAGCAGCGCCAGCCAGCCAGCAGAGAGAAAGTCAAGTACCC 350
481 oGluIleIlePheLeuGlyThrGlySerAlaIleProMetLysIleArgA 498
351 AGAAATCATCTTCTTGGAAACAGGGTCTGCCATCCGATCCGATTCGAA 400
498 snValSerAlaThrLeuValAsnIleSerProAspThrSerLeuLeuLeu 514
401 ATGTCACTGCCACACTGTCAACATTAAGCCCGACAGCTCTGTGCTACTG 450
515 AspCysGlyGluGlyThrPheGlyGlnLeuCysArgHisTyrGlyAspGL 531
451 GACTGTGGTCAAGGCACATTTGGCAGCTGTGCCGTCAATACGAGACCA 500
531 nValAspArgValLeuGlyThrLeuAlaAlaValPheValSerHisLeuH 548
501 GGTGGACAGGGTCTGGGACCCCTGGCTGTGTGTGTGTGTGTGTGTGTGT 550
548 isAlaAspHisHisThrGlyLeuProSerIleLeuLeuGlnArgGluArg 564
551 ACCGACATCACCACAGCGCTTGCCTCAAGTATCTTGTGCAAGAGAACG 600
565 AlaLeuAlaSerLeuGlyLysProLeuHisProLeuValValAlaPr 581
601 CGCTTGGCATCTTGGGAAAGCGCTTACCCCTTGTGTGTGTGTGTGTGTG 647
581 oAsnGlnLeuLys 585
648 CCAACAGTCAAG 660
```

seq_name: gb_est70:BE382353

seq_documentation_block:

LOCUS BE382353 692 bp mRNA EST 21-JUL-2000
DEFINITION 601298656F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:3629028 5',
mRNA sequence.
ACCESSION BE382353
VERSION BE382353.1 GI:9327718
KEYWORDS EST.
SOURCE human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 692)

NIH-MGC <http://www.ncbi.nlm.nih.gov/MGC/>.

AUTHORS

National Institutes of Health, Mammalian Gene Collection (MGC)

TITLE

Unpublished (1999)

JOURNAL

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Ling Hong/Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov

Plate: L10M314 row: O column: 13

High quality sequence stop: 600.

Location/Qualifiers

1..692

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:3629028"

/clone_lib="NIH_MGC_19"

/tissue_type="neuroblastoma"

/lab_host="DH10B (phage-resistant)"

/note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:

EcoRI; cDNA made by oligo-dT priming. Directionally

cloned into EcoRI/XhoI sites using the following 5'

adaptor: GGCAGCAG(G).. Library constructed by Ling Hong

in the laboratory of Gerald M. Rubin (University of

California, Berkeley) using ZAP-cDNA synthesis kit

(Stratagene) and Superscript II RT (Life Technologies).

Note: this is a NIH_MGC Library."

BASE COUNT 152 a 193 c 197 g 150 t

ORIGIN

alignment_scores:

Quality: 1066.00 Length: 224

Ratio: 4.912 Gaps: 3

Percent Similarity: 96.875 Percent Identity: 95.089

alignment_block:

US-09-434-382-2 x BE382353 ..

Align seg 1/1 to: BE382353 from: 1 to: 692

493 ProMetLysIleArgAsnValSerAlaThrLeuValAsnIleSerProAs 509

3 CCGATGAAGATTGCAATGTCAGTGGCCACACTTGTCAACATAAGCCCCGA 52

509 pThrSerLeuLeuAspCysGlyGluGlyThrPheGlyGlnLeuCysA 526

53 CAGCTCTCTGTACTGGACTGTGTGAGGCGACATTGGGCGACTGTGCC 102

526 rGHisTyrGlyAspGlnValAspArgValLeuGlyThrLeuAlaVal 542

103 GTCATTCAGGAGACAGGTCGAGGTCCTGGGACCCCTGGCTGTGTG 152

543 PheValSerHisLeuHisAlaAspHisHisThrGlyLeuProSerIleLe 559

153 TTTTGTGTCCACCTGCAGCGAGATCACACACGGGCTTGCAAGTATCTT 202

559 uLeuGlnArgGluArgAlaLeuAlaSerLeuGlyLysProLeuHisProL 576

203 GCTGCGAGAGACGCGCTTGGCATCTTTGGGAAGCGCGCTTCACCCCT 252

576 euLeuValValAlaProAsnGlnLeuLysAlaTrpLeuGlnTyrHis 592

253 TGCTGTGTGTGTCGCCCAACACAGCTCAAGCCCTGGCTCCAGCAGTACC 302

593 AsnGlnCysGlnGluValLeuHisHisIleSerMetIleProAlaLysC 609

```

|||||
303 AACCAAGTCCAGGAGGTCCTGCACCACATCAGTATTCCTGCGCAATG 352
|||||
609 sLeuGlnGluGlyAlaGluIleSerSerProAlaValGluArgLeuIleS 626
|||||
353 CCTTCAGGAAGGGCTGAGATCTCCAGTCCCTGCAGTGGAAAGATTGATCA 402
|||||
626 erSerLeuLeuArgThrCysAspLeuGluGluPheGlnThrCysLeuVal 642
|||||
403 GTTCGCTGTTGCGAACATGTGATTGGAAGAGTTTCAGACCTGTCTGGTG 452
|||||
643 ArgHisCysLysHisAlaPheGlyCysAlaLeuValHisThrSerGlyTr 659
|||||
453 CGGCACCTGCAAGCATGCGTTTGGGTGCTGCTGTCACACTCTGGCTG 502
|||||
659 pLysValValTyrSerGlyAspThrMetProCysGluAlaLeuValArgM 676
|||||
503 GAAAGTGTCTATTCCGGGGACACCATGCTGCTGCGAGGCTCTGGTCCGA 552
|||||
676 etGlyLysAspAlaThrLeuLeuIleHisGluAlaThrLeuGluAspGly 692
|||||
553 TGGGAAAGATGCCACCTCTCTGATACATGAAGCCACCTG.GAAGATGCT 601
|||||
693 LeuGluGluAlaValAlaGluLysThrHisSerThr.ThrSerGlnAlaI 709
|||||
602 TTGGCAGAGGAAGCGGTGGAAGA...CACACGACACACGTCCTCCAAACA 647
|||||
709 leSerValGlyMetArgMet 715
|||||
648 ..TCACGTGGGATCGGGATA 665

```

seq_name: gb_est69:BE260626

seq_documentation_block: 938 bp mRNA EST 13-JUL-2000
LOCUS BE260626 601146116F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:3161691 5',
DEFINITION mRNA sequence.

ACCESSION BE260626
VERSION BE260626.1 GI:9132065
KEYWORDS EST.
SOURCE human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 938)
NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov

Tissue Procurement: ATCC

CDNA Library preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLN at: image.llnl.gov
Plate: L1CM119 row: o column: 04
High quality sequence stop: 621.
Location/Qualifiers

FEATURES

source

```

1..938
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:3161691"
/clone_lib="NIH_MGC_19"
/tissue_type="neuroblastoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: brain; Vector: pOTB7; Site: 1: xhoI; Site 2:  

EcoRI; cDNA made by oligo-dT priming. Directionally  

cloned into EcoRI/XhoI sites using the following 5'  

adaptor: GGACCGAG(G). Library constructed by Ling Hong  

in the laboratory of Gerald M. Rubin (University of  

California, Berkeley) using ZAP-cDNA synthesis kit
```

(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
BASE COUNT 190 a 263 c 283 g 202 t
ORIGIN

alignment_scores:

Quality: 1005.00 Length: 290
Ratio: 4.102 Gaps: 10
Percent Similarity: 84.483 Percent Identity: 75.862

alignment_block:

US-09-434-382-2 x BE260626 ..

Align seg 1/1 to: BE260626 from: 1 to: 938

```

261 GluMetGlyLeuProValGlyThrAlaAlaIleAlaProIleIleAlaAl 277
|||||
1 GAGATGCTTTTCCCGTCTCTGCAGCAGCAATGATGGAGCGATGGCAGC 50
|||||
277 aValLysAspGlyLysSerIleThrHisGluGlyArgGluIleLeuAlaG 294
|||||
51 TGTCCCAACTGGAGAGCATCACTCATGAAGGAGAGAGATTTTGGCTG 100
|||||
294 luGluLeuCysThrProProAspProGlyAlaAlaPheValValGlu 310
|||||
101 AAGAGCTGTACTCTCCAGATCTCTGCTGTC. TTTGTGTGTGTAGAA 149
|||||
311 CysProAspGluSerPheIleGlnProIleCysGluAsnAlaThrPheGl 327
|||||
150 TGTCCAGATGAAGCTTCATTCAACCCCATCTGTGAGAATGCCACCTTCA 199
|||||
327 nArgTyrGlnGlyLysAlaAspAlaProValAlaLeuValHisMetA 344
|||||
200 GAGGTACCAAGAAAGGAGATGCCCGCTGGCTTGGTGGTTACATGG 249
|||||
344 laProAlaSerValLeuValAspSerArgTyrGlnGlnTrpMetGluArg 360
|||||
250 CCCAGCATCTGTCTTGTGGACAGCAGGTACCAGCAGTGGATGGAGAG 299
|||||
361 PheGlyProAspThrGlnHisLeuValLeuAsnGluAsnCysAlaSerVa 377
|||||
300 TTTGGGCTGACACCCACACTTGTGCTGAATGAGAATGTGCTCAGT 349
|||||
377 lHisAsnLeuArgSerHisLysIleGlnThrGlnLeuAsnLeuIleHisP 394
|||||
350 TCACACCTTCGCGCCACACAGATTCAACCCAGCTCAACCTCATCCACC 399
|||||
394 roAspIlePheProLeuLeuThrSerPheArgCysLysLysGluGlyPro 410
|||||
400 CGGACATCTTCCCGCTGCTCACCAGTTTCCGCTGTGAAGAGAGGGCCCC 449
|||||
411 ThrLeuSerValProMetValGlnGlyCysLeuLeuLysTyrGlnLe 427
|||||
450 ACCCTCAGTGTGCCCATGGTTTCAGGGTGAATGCCCTCAAGTACCAGCT 499
|||||
427 uArgProArgArgGluTrpGlnArgAspAlaIleIleThrCysAsnProG 444
|||||
500 CCGTCCCGAGGAGGTGGCAGAGGGATGCCATTATTACTTGCATCTCTG 549
|||||
444 luGluPheIleValAlaLeuGlnLeu.ProAsnPheGlnGlnSerVa 460
|||||
550 AGGAATTCATAGTTGAGGCGCTGCAGCTTCCCAACTTCCAGCAGAGCGT 599
|||||
460 lGlnGluTyrArgSerAlaGlnAspGlyProAlaProAlaGluLysA 477
|||||
600 CGCGAGT...ACAGAGAGAGTGGCGGAGGGCCAGCCCCCGCAGAGAA...A 643
|||||
477 rgSerGlnTyrProGluIlePheLeuGlyThrGlySerAlaIlePro 493
|||||
644 GAAGTCGTTTCCCGGAT...CTCTCTCTG...GACGGGTCTGGCATCCCG 687
|||||
494 MetLysIleArgAsnValSerAlaThrLeuValAsnIleSerPro..... 508

```

```

688 GTGAAGATTGCA.....ATGTCGTCGCCCTGTTCA 719
509 .AspThrSerLeuLeuLeuAspCysGlyGluGlyThrPheGlyGlnLeu 525
720 TAAGCCGCCGCTCTGTACTGTGTTGCGGCTTTGGCG.....T 763
525 ysArgHis....TyrGlyAspGlnValAspArgValLeu.....Gly 537
764 GCCGTACGACAGAGGGGGGCTGGCCCGCTGTGTGTTGTCGCCGGGG 813
538 ThrLeuAlaAlaValPhe 543
814 ACAACGGGGCGGCTTTT 831

```

seq_name: gb_est75:BE742908

seq_documentation_block: 677 bp mRNA EST 15-SEP-2000
 LOCUS BE742908 601574609F1 NIH_MGC_9 Homo sapiens cDNA clone IMAGE:3835658 5',
 DEFINITION mRNA sequence.

ACCESSION BE742908
 VERSION BE742908.1 GI:10156900
 KEYWORDS EST

SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 677)

AUTHORS NIH-MGC <http://www.ncbi.nlm.nih.gov/MGC/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

Tissue Procurement: DCTD/DTP

cDNA Library Preparation: Ling Hong/Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov

Plate: LLCMS18 row: a column: 03

High quality sequence stop: 672.

Location/Qualifiers

SOURCE

1. 677

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:3835658"

/clone_lib="NIH_MGC_9"

/tissue_type="adenocarcinoma cell line"

/lab_host="DH10B (phage-resistant)"

/note="Organ: ovary; Vector: pOTB7; Site_1: XhoI; Site_2:

cloned into EcoRI/XhoI sites using the following 5'

adaptor: GGCAAGAG(G). Size-selected >500bp for average

insert size 1.8kb. Library constructed by Ling Hong in

the laboratory of Gerald M. Rubin (University of

California, Berkeley) using ZAP-cDNA synthesis kit

(Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 160 a 173 c 217 g 127 t

ORIGIN

alignment_scores:

Quality: 1000.50 Length: 226

Ratio: 4.611 Gaps: 3

Percent Similarity: 96.018 Percent Identity: 95.133

alignment_block:

US-09-434-382-2 x BE742908

Align seg 1/1 to: BE742908 from: 1 to: 677

```

603 SerMetIleProAlaLysCysLeuGlnGluGlyAlaGluIleSerSerPr 619
1 AGTATGATTCTCCCAATGCCTTCAGGAAGGGGCTGAGATCTCCAGTCC 50
619 oAlaValGluArgLeuIleSerSerLeuLeuArgThrCysAspLeuGluG 636
51 TCAGTGGAAAGATTGATCAGTTTCGCTGTTCGGAACATGTGATTGGAAG 100
636 luPheGlnThrCysLeuValArgHisCysLysHisAlaPheGlyCysAla 652
101 AGTTTCAGACCTGTCTGTGTCGCGCAGCTCAAGCATGCGTTTGGCTGG 150
653 LeuValHisThrSerGlyTrpLysValValTyrSerGlyAspThrMetPr 669
151 CTGGTGCACACCTCTGGCTGGAAGTGGTCTATTTCGGGGACACCATGCC 200
669 oCysGluAlaLeuValArgMetGlyLysAspAlaThrLeuLeuHisG 686
201 CTGCGAGGCTCTGGTCCGATGGGAAAGATGCCACCTCCTCATACATG 250
686 luAlaThrLeuGluAspGlyLeuGluGluAlaValGluLysThrHis 702
251 AAGCCACCTCGGAAGATGGTTG.GAAGAGAGACAGCTGGAAAGACACAC 299
703 SerThrThrSerGlnAlaIleSerValGlyMetArgMetAsnAlaGluPh 719
300 AGCACAACTCCCAAGCCATCAGCGTGGGATGCGGATGAACGCGGAGTT 349
719 erIleMetLeuAsnHisPheSerGlnArgTyrAlaLysValProLeuPheS 736
350 CATTAATGCTGAACCACTTCAGCCAGCGCTATGCCAAGGTCCCTCTTCA 399
736 exProAsnPheSerGluLysValGlyValAlaPheAspHisMetLysVal 752
400 GCCCAACTTCAGCGAAGTGGGAGTTGCCCTTGACCACATGAAGGTC 449
753 CysPheGlyAspPheProThrMetProLysLeuIleProProLeuLysAl 769
450 TGCTT.GGAGACTTTTCAACAATGCCCAAGCTGATTCCTCCCTCAAGGC 498
769 a.LeuPheAlaGlyAspIleGluGluMetClnuArgArgGluLysArg 785
499 CCGTGTGTTGCTGGCGACATCGAGGAGATGGAGAGCGGCAGGAGAGCGG 548
786 GluLeuArgGlnValArgAlaAlaLeuLeuSerArgGluLeuAlaGly.. 801
549 GAGCTCGCGCAGGTGCGGGGGCC.....TCCTGTCCAGGAGCTGGCAG 592
802 .GlyLeuGluAspGlyGluProGlnLysArgAlaHisThrGluGluP 818
593 GCGGCTGGAGGATGGGAGCCTCAGCAGAAGCGGGC.CACACAGAGGAGC 641
818 roGlnAlaLysLysValArgAlaGln 826
642 CACAGCCCAAGAAGGTTCAGAGCCCGAG 667

```

